Author:
Sheng, Jie

Title:
Managing Big Data from the Crowd

Strategic Firm Engagement with Online Social Interactions

Managing Big Data from the Crowd:
Strategic Firm Engagement with Online Social Interactions

Jie SHENG

A dissertation submitted to the University of Bristol
in accordance with the requirements for award of the degree of
Doctor in Philosophy
in the Faculty of Social Sciences and Law
School of Economics Finance and Management

June 2018

Word Count: 71,456

ABSTRACT

In today's digital economy, information sharing has become common practice and significantly influences individuals' behaviours and preferences. The interactive and participative environment fosters customer engagement in voicing and communicating in the virtual network. The sheer amount of user-generated content from online social interactions offers intriguing opportunities for businesses to develop sustainable competitive advantages; yet, how firms can create value by managing and capitalising on crowd voices remains an under-explored facet of big data research. This thesis discusses strategies for firms to engage in the online social interaction network so as to improve performance and achieve competitive advantages. The developed holistic framework for strategic firm engagement articulates three distinct but non-mutually exclusive roles of firms in the online communication network: observer, participant and strategic leader. Correspondingly, three studies are designed to examine business impacts of these firm engagement roles using a large-scale data set of over 800,000 online customer reviews and over 360,000 online managerial responses of London hotels. The first study investigates the observer role and validates an analytical strategy for mining customers' textual reviews and exploiting the discovered knowledge to improve service quality. The second study considers the participant role and explores how firms respond to customer reviews and the efficacy of different response styles in future rating improvement. The third study examines the strategic leader role by testing the effects of firms being present and active online in stimulating customer engagement behaviour. Findings from the empirical studies demonstrate the strategic value of firm engagement in the online social interaction network. This thesis contributes to big data research, strategy and marketing literature in terms of strategising big data from the crowd by developing data-driven strategies. It also offers practical insights into strategic planning for businesses engaging in online social interactions.

ACKNOWLEDGEMENT

I would like to extend thanks to many people for their immense support during my PhD pursuit.

First, I wish to express my deep gratitude to my supervisors, Xiaojun Wang and Joseph Amankwah-Amoah. I am extremely lucky to have them as my PhD supervisors. They have been tremendously supportive by guiding me academically and emotionally through the road to finish this thesis and I thank them wholeheartedly for all their contributions of time and ideas and their constant faith in my work. I truly appreciate their effort not only in making my PhD experience productive and stimulating, but also in enriching my academic career. The enthusiasm they have for research is also motivational for me and I am thankful for the excellent role models they have provided as dedicated and successful researchers.

My sincere thanks go to the School of Economics Finance and Management for giving me the opportunity to carry out my research and for funding my doctoral study. I would like to thank senior researchers in our School, in particular, Department of Management, for their academic support. Special thanks to Tamer Elshandidy, Ahmed Mostafa, and Robin Klimecki for being my upgrade panel members and providing valuable feedback on a preliminary version of this thesis. Huge thanks to my examiners, Agnes Nairn and Mike Tse, for examining this thesis and making the viva voce intellectually stimulating and enjoyable. I am indebted to David Dugdale for his help and encouragement to me to embark on the academic path. I am also thankful for Leroy White who offered me the PhD position at the first place, and Thierry Viale who agreed to be a member of the supervision team. Also thank you Stephen Pearce and Rutvica Andrijasevic for the great teaching experience that inspires my research in many ways.

Heartfelt thanks to all my friends, in China and in the UK, who have always been encouraging and supportive. Some special words of gratitude go to all my PhD colleagues, past and present, for their presence and sharing moments of excitement and anxiety. In particular, a warm word for my colleagues and friends, Izah and Anisha, with whom I have had the best tea breaks in my life. Thank you girls for always being there for me through this three-and-a-half year journey. I would also like to thank Dale, Xiaoting, Claire, Alice, and many others for their kind help at different stages of my PhD study. I am grateful for many other people in room 0C3, and I have very fond memories of my time there.

Lastly, I am hugely appreciative of my family for all their love and encouragement. A very special word of thanks goes to my parents who have always supported me unconditionally in all my pursuits and been showing how proud they are of me. I would like to thank my grandparents for their continuous support and love. They have given me the extra motivation to finish the degree. For my loving and supportive husband, Hao, thank you for keeping things going and being a constant source of strength and inspiration for me. My PhD time was enjoyable because of him. I dedicate this thesis to my family.

AUTHORS' DECLARATION

I declare that the work in this dissertation was carried out in accordance with the requirements of the University's *Regulations and Code of Practice for Research Degree Programmes* and that it has not been submitted for any other academic award. Except where indicated by specific reference in the text, the work is the candidate's own work. Work done in collaboration with, or with the assistance of, others, is indicated as such. Any views expressed in the dissertation are those of the author.

SIGNED: .... DATE: ....

STATEMENT OF CONTRIBUTIONS

Part of Chapter 2 "Literature Review" was published at *International Journal of Production Economics*, with joint authorship of my PhD supervisors Dr Joseph Amankwah-Amoah and Dr Xiaojun Wang, as:

Sheng, J., Amankwah-Amoah, J., & Wang, X. (2017). A multidisciplinary perspective of big data in management research. *International Journal of Production Economi*cs, 191, 97-112.

The statement of contributions of the authors is as follows and this declaration is jointly authorised by the signature of the parties below:

Name of the author	**Contributions**	**Signature**
Jie Sheng	Conception, data collection, data analysis, data interpretation, writing, and revision.	Jie Sheng
Joseph Amankwah-Amoah	Supervision and editing	[illegible]
Xiaojun Wang	Supervision and editing	X Wang

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

CHAPTER 1

INTRODUCTION

1.1 Research Background

'Big data' has emerged as a prominent buzzword in recent years. In the age of the Internet of Things, data is everywhere, with significant scale escalation and scope expansion. The term 'big data' describes such a tremendous explosion of information as an extremely large data pool characterised by significant volume, variety, velocity, veracity, variability and value (Gandomi & Haider, 2015; Jin, Wah, Cheng, & Wang, 2015; Katal, Wazid, & Goudar, 2013). There is growing recognition of the big data significance across industries and sectors (Gandomi & Haider, 2015). One reason that big data attracts massive attention is the hidden value in enhancing productivity and creating social surplus (Manyika, Chui, Brown, Bughin, Dobbs, Roxburgh, & Byers, 2011). Such a rich diversity of big data has been regarded as "enterprise assets" which can "yield actionable business insights" (Russom, 2011, p. 9). Such potential values stem from enhanced accuracy in decision-making processes, in which big data paves the way for decision makers to act in a timely and data-driven manner based on solid data evidence rather than intuition (Beath, Becerra-Fernandez, Ross, & Short, 2012; Fosso Wamba, Akter, Edwards, Chopin, & Gnanzou, 2015; McAfee & Brynjolfsson, 2012).

Meanwhile, the complexity inherent in big data poses great challenges in creating and capturing value. To transform the massive data into knowledge and meaningful insights, business analytics plays an essential role (Chen, Chiang, & Storey, 2012). Business analytics is concerned with methods and techniques for dealing with data to assess past or real-time business performance (Davenport, 2006). Facing the big data challenge, technological breakthroughs and innovation are in urgent need of enabling efficient collection, storage and process of the enormous amount of structured, unstructured and semi-structured data coming from multiple sources (Davenport, Barth, & Bean, 2012; Turban, Sharda, Aronson, & King, 2008; Watson & Wixom, 2007). By combining massive data sets and advanced analytics techniques, big data analytics has now become a trendy practice in business intelligence (Russom, 2011). With extensive data collected and interpreted, companies are able to investigate the specifically detailed aspects of business activities and to identify the ongoing position of the business (LaValle, Lesser, Shockley, Hopkins, & Kruschwitz, 2011) so as to establish competitive advantages (Fosso Wamba, Gunasekaran, Akter, Ren, Dubey, & Childe, 2017).

In light of the big data phenomenon and the changing business environment with a growing interest in big data analytics, this thesis posits that organisations could enhance their adaptability and dynamics by developing data-driven strategies to grasp the opportunity to create sustained value from big data. Situated in the context of big data, the present author performs big data analytics to explore how businesses can strategise about big data to enhance competitiveness and managerial effectiveness and thereby accrue superior benefits in the marketplace.

1.2 Research Scope and Motivation

One type of big data is user-generated content (UGC), which is created by Internet users with the purpose of expressing themselves publicly and sharing information in the virtual community (Ayeh, Au, & Law, 2013). UGC can take many forms, such as online comments, blogs, video/audio files, images, and posts on discussion forums. Enabled by technological advancements, there has been a proliferation of UGC in recent years that has been created and exchanged in a more rapid, transparent, and frequent manner (Hennig-Thurau, Malthouse, Friege, Gensler, Lobschat, Rangaswamy, & Skiera, 2010). Individual opinions can be directly delivered to a wide range of people and organisations via the Internet (Hennig-Thurau, Gwinner, Walsh, & Gremler, 2004). This new mechanism of information exchange and diffusion has changed the business landscape. The networked world creates a novel opportunity for firms to steer business towards a sustainable and dynamic competitiveness. For instance, UGC contains rich information about the user experience and opinions towards certain products, services, and brands, which is highly informative about markets (Cachia, Compañó, & Da Costa, 2007). Higher efficiency and effectiveness of management practices and strategic decisions are potentially achievable through co-creating value with customers and mining user-created data (Prahalad & Ramaswamy, 2013).

Along with the spread of UGC, there are augmented online social interactions (OSIs) among end users. Godes et al. (2005) define social interaction as “an action or actions that (a) is taken by an individual not actively engaged in selling the product or service and that (b) impacts others' expected utility for that product or service" (p. 416–417). This broad term captures a wide range of contexts where interpersonal communication takes place. In the business and management field, social connections often refer to information exchange, opinion sharing, chit-chat or gossip among consumers (Berger, 2014). In a traditional format, the communication takes place between one communicator who is not involved in selling a product/service and another person concerning the offerings or the brand (Arndt, 1967). With the advent of the Internet, the "one-to-one and face-to-face exchange of information about a product or service" (Godes et al., 2005, p. 416) has been extended to the virtual Internet-connected world and becomes one-to-many and anonymous.

The emerging OSIs in the virtual community are becoming increasingly important in digital economies and have far-reaching implications for businesses. As illustrated by Berger (2014), interpersonal communications have five key functions: impression management, emotion regulation, information acquisition, social bonding, and persuading others. Motivated by these intrinsic and/or extrinsic factors, individuals participate in OSIs by generating and/or receiving information (De Matos & Rossi, 2008; Godes et al., 2005). For example, electronic word-of-mouth (eWOM) is "any positive or negative statement made by potential, actual, or former customers about a product or company, which is made available to a multitude of people and institutions via the Internet" (Hennig-Thurau et al., 2004, p. 39). It is viewed as a credible information source in a way that it is offered by fellow consumers and is often supplementary information in addition to the seller-side stories. Extensive research has suggested that eWOM has substantial impacts on consumers' perceptions and behaviour (e.g., Chevalier & Mayzlin, 2006; Hennig-Thurau, Wiertz, & Feldhaus, 2015; Sonnier, McAlister, & Rutz, 2011) and, thereby, on business performance (Cascio, O'Donnell, Bayer, Tinney Jr, & Falk, 2015; Cheung, Xiao, & Liu, 2014; Peters, Chen, Kaplan, Ognibeni, & Pauwels, 2013).

In view of the proliferation of crowd voice and its potential impacts, understanding and managing online interpersonal communications are critical to the development and validation of business strategies. Over the last decade, it has been recognised that firms have incentives to manage OSIs, and firm-level engagement in OSIs has drawn interest from both academia and industry (Evans, 2010; Yadav & Pavlou, 2014). Firm engagement, also referred to as business engagement, emphasises that businesses are not only aware of the importance of OSIs but are also taking a role in the OSI network by engaging in online communications (Zhang, Jansen, & Chowdhury, 2011). In fact, with advanced business intelligence capability, companies are increasingly managing and engaging in OSIs so as to compete in the digital era (Godes et al., 2005). For example, firms monitor customers' interactions for the purpose of discovering knowledge and innovating new ideas (e.g., Nambisan & Baron, 2007). Social meida is employed by businesses as a marketing tool to generate and deliver messages to targeted audiences (e.g., Gu & Ye, 2014; Porter & Donthu, 2008) and to enhance their online reputation (e.g., Dijkmans, Kerkhof, & Beukeboom, 2015). These examples reveal that active management of OSIs has promising value in exerting influence in the network and contributing to business decision-making, thereby improving management efficiency and profitability.

Nonetheless, according to Godes et al. (2005) and Kumar and Pansari (2016), firm engagement appears to be an under-explored topic in the current state of knowledge. Extant research has mainly focused on interpreting OSIs among customers, attempting to discern customer engagement behaviour in the OSI network (e.g., Hennig-Thurau et al., 2004; Van Doorn, Lemon, Mittal, Nass, Pick, Pirner, & Verhoef, 2010). While it is valuable to uncover the importance of OSIs from the customer perspective, the OSI network also involves other actors, such as firm managers and employees, regulators, and the public. They play different roles in online communication and their engagement may also have distinct influences on businesses. Kumar and Pansari (2016) point out that the focus has now shifted, with 'engagement' reflecting the connectedness among customers, employees and the firm. Nevertheless, there is a lack of a holistic framework demonstrating how firms engage and what positions firms hold in the OSI network. It also remains unclear what the consequences of firm engagement are and the extent of these potential influences on customers and businesses. Furthermore, limited empirical evidence has been documented to suggest how businesses develop effective strategies for acting on and engaging in OSIs to achieve strategic value. Accordingly, this research is conducted in the context of big data and particularly focuses on the big data generated from eWOM communications among the crowds of consumers in their OSIs. By analysing these data using advanced analytics techniques, this research explores how businesses capitalise on crowd voices by acting on OSIs.

1.3 Research Questions and Objectives

This thesis attempts to fill the identified research gap by exploring OSIs from a business perspective. In particular, this scholarly work aims to discuss firm engagement in OSIs at the strategic level and empirically examine the potential impacts and effectiveness of firm engagement strategies. To fulfil the research aim, there are four key questions that guide this research:

Research Questions (RQs)

RQ 1: What roles can firms play in the OSI network?

RQ 2: What are the business impacts of firms managing and engaging in OSIs?

RQ 3: To what extent can firm engagement in OSIs create benefits for businesses?

RQ 4: How can firms strategically engage in OSIs to achieve favourable outcomes?

The above questions pertaining to firm engagement in OSIs underlie the objectives of this research:

Research Objectives (ROs)

RO 1: To clarify the tenets of firm engagement by identifying firms' roles and activities in managing and engaging in OSIs.

RO 2: To demonstrate the relationship between firm engagement in OSIs and business performance by analysing the consequences of firms playing distinct roles in the OSI network.

RO 3: To identify the mechanism and to evaluate the efficacy of engaging and exploiting OSIs in creating strategic values.

RO 4: To develop effective strategies for firm engagement in OSIs and to offer practical insights into data-driven decision-making.

1.4 Research Framework

This section explains the conceptual framework used to organise this thesis (see Figure 1-1). The framework is inspired by Godes et al. (2005), who have introduced four generic implementation strategies in managing social interactions.[1] The proposed framework depicts the positions and scope of firms' different roles in the OSI network, thus responding to the first research question regarding the roles of firms and the functionality of each role in managing OSIs.

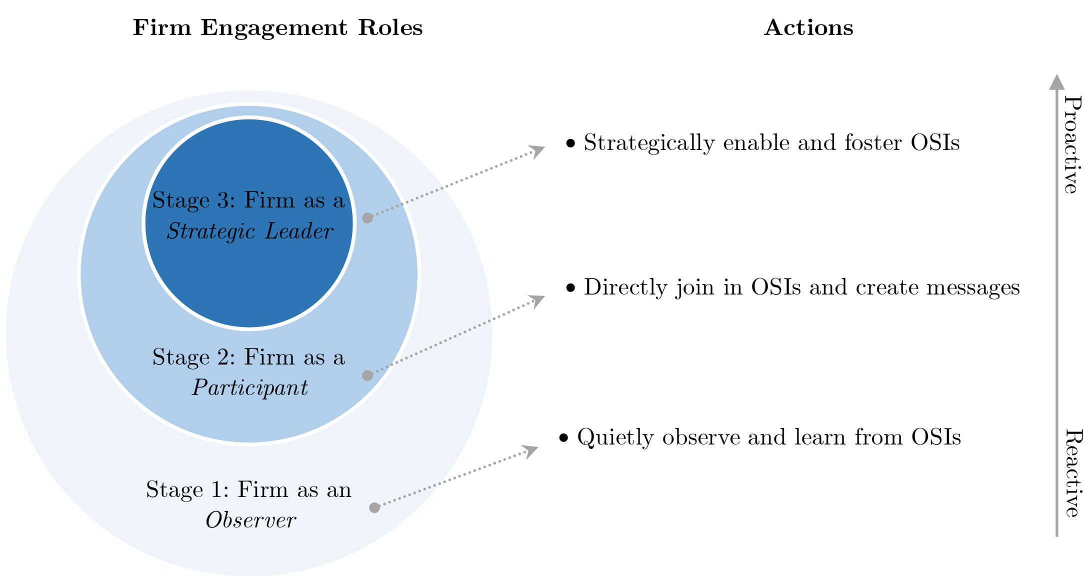


Figure 1-1. Firm engagement in the OSIs network

A working definition of firm engagement in this research is: *firms' strategic intention and behavioural manifestation of it, motivated by multifaceted incentives, to engage in the OSI network and to act on OSIs among consumers.* Firms' active management of social interactions can be realised in various ways in the OSI network. In this research, firm-level engagement is reflected by businesses taking three different but non-mutually exclusive roles—namely, from reactive to proactive, *observer*, *participant*, and *strategic leader*—whereby firms engage in OSIs by learning, joining and leading online communications.

[1] Godes et al. (2005) discuss this issue by identifying four generic strategies—firms act as an observer, a moderator, a mediator, or a participant—to manage social interactions. The observer position refers to collecting information and learning about the ecosystem from the conversations. A firm can also act as a moderator to foster social interactions by establishing online community and customer recommendation programmes. For the mediator strategy, information is under control of a firm which spreads information itself. The most aggressive action is participating, which involves creating word-of-mouth and strategic manipulation of online communications. In contrast to their work, the current research defines three roles that firms play in the OSI network from a strategic perspective based on the degree of proactivity in business online engagement.

Observer. As an observer, firms collect OSI information (e.g., tracking customer engagement behaviour, recording interpersonal communication messages) created by customers and learn from it by analysing the data. This is a relatively silent role in which firms hide behind the screen and extract information that is potentially useful for the business. The key concerns for the role of an observer are the purpose of observing, the methods for analysing massive data and the assessment of learning outcomes. The intentions to manage and engage via observing and learning from OSIs vary with business goals; nevertheless, this process can be potentially used to perform online evaluation of firm performance (e.g., brand equity, market position, operation efficiency), improve product or service quality (e.g., product defect discovery, service design), innovate new products, and dynamically adjust strategies (e.g., competitive pricing and inventory management).

Participant. Going beyond the silent role, a more active option is to become a participant, directly joining online communications and creating online messages.[2] In this position, firms establish conversations with customers, and their voices can be heard by consumers and the public. For example, firms may respond to customers' praise or complaints, deal with customer queries, and provide additional information upon request on the online platform. This role brings about interactions between firms and customers as well as interactivity across the online communication network, and the firm-to-customer messages constitute an element of eWOM from the OSIs. The key questions concerned with the participant role are how firms participate (e.g., frequency of posting, information and attitudes conveyed in the messages) and how firm–customer conversations exert influence within the OSI network.

Strategic leader. The most proactive role in the proposed framework is that of the firm as a strategic leader, taking a step beyond silent observation and reactive participation. In this role, a firm leads OSIs on its own initiative. For example, firms can act as a facilitator to foster and enable consumers' OSIs by establishing an online communication platform or community. Firms may also monitor online information diffusion and deliberately circulate messages to certain groups of consumers (e.g., social media marketing campaigns) or even manipulate opinions in

[2] Unlike Godes et al. (2005), who view participation as the most aggressive action, this research argues that such participation is, to some extent, reactive in nature. In the participation setting, OSIs among consumers happen first, and firms react to the eWOM by posting corresponding messages. It is not the firm that initiates and leads the online conversations. Nevertheless, a firm's choice to participate and how to participate are more proactive. This links to the most proactive role of strategic leaders.

online forums (e.g., Dellarocas, 2006) to ensure that relevant OSI information is favourable to firms. Hence, this strategy involves firms acting strategically and dynamically in a network and is mainly implemented to encourage and monitor OSIs. The central issues regarding the strategic leader role are the potential benefits that firms could obtain from devoting efforts to actively managing OSIs and how effective the strategy is in influencing consumer behaviour in a network.

1.5 Research Methodology

Research methodology refers to a systematic plan for addressing a research problem (Kothari, 2004). This section clarifies the methodology of this research, including the type of research, research design and specific methods for data collection and data analysis.

1.5.1 Type of research

This subsection defines this research from both philosophical and methodological viewpoints (see a summary in Table 1-1).

Table 1-1. Research types

Research	Types of research	The current research
Paradigm	Functionalist, Interpretive, Radical humanist, Radical structuralist	Functionalist
Philosophical position	Ontology: Objectivism, Subjectivism Epistemology: Positivism, Realism, Interpretivism Axiology: Value-free or not	Objectivism Positivism Value-free
Approach/Logic	Deductive, Inductive	Deductive and Inductive
Purpose	Exploratory, Descriptive, Explanatory	Explanatory and Exploratory
Strategy	Experiment, Survey, Case study etc.	Big data analytics
Methodological choice	Quantitative, Qualitative, Mixed method	Mixed methods
Time horizons	Cross-sectional, Longitudinal	Cross-sectional and Longitudinal

Research paradigm. This research belongs to the functionalist paradigm, which combines objectivist and regulatory dimensions and aims to provide a rational explanation and practical solutions to a practical problem (Saunders, Lewis, & Thornhill, 2016). The current research explores how firms can strategise in relation to data created in OSIs by managing and engaging in the OSI network. The studies focus on developing strategies, assessing the effectiveness of these strategies, and offering recommendations to businesses for making the strategies more effective.

Research philosophy. The philosophical stance can be viewed from three perspectives: ontological, epistemological and axiological (Saunders et al., 2016). In terms of ontology, this research reflects the philosophy of objectivism, in which the nature of reality is external, objective and independent of social actors. In terms of epistemology, this research reflects the philosophy of positivism, which focuses on causality and generating laws or principles based on data and facts from observable phenomena. In terms of axiology, this research is undertaken without value influence and the researcher maintains an independent and objective standpoint. In this regard, the current research draws on the big data phenomenon and collects data of online UGC to study the observable reality and to develop law-like generalisations, which corresponds with an objectivist, positivist and value-free philosophical approach.

Research approach/logic. This thesis combines an inductive approach (Chapter 3) and a deductive approach (Chapters 4 and 5). Inductive research formulates a theory based on understanding and analysing qualitative data. This applies to Study I (Chapter 3), in which the researcher analyses a large set of textual data to construct topic models and interprets the topics based on a review of relevant literature. In contrast, deductive research develops hypotheses based on theories and performs a rigorous test to prove or disprove the theories. In this thesis, Study II (Chapter 4) and Study III (Chapter 5) employ the deductive approach to explain causal relationships between variables with a highly structured model and numeric data (including quantified textual data).

Research purpose. The purpose of research is determined by the research questions and objectives. This thesis presents an exploratory study (Study I, Chapter 3) and two explanatory studies (Study II in Chapter 4 and Study III in Chapter 5). An exploratory study seeks to understand a phenomenon and to develop new insights, while an explanatory study investigates a problem by establishing and explaining causality in that situation (Saunders et al., 2016). According to this classification, the first study is an exploratory study aimed at understanding consumers' concerns in terms of service quality by analysing textual content of UGC and interpreting the thematic information emerging from the data against the current knowledge in the literature. The second and third studies are explanatory studies that hypothesise and test the causal relationships between firm actions and consumer behaviour.

Research strategy. A research strategy helps to achieve the research aim and to answer research questions. The choice of a research strategy is informed by the research objectives and the types of research questions (Sekaran & Bougie, 2016). Various research strategies are commonly used in social science studies, such as experiment, interview, survey, case study, action research, and archival research. This thesis employs empirical research, applying a big data analytics approach to a large set of secondary data collected from an online community platform. Big data analytics can be defined as operating advanced analytics techniques on big data so as to enable the data to "speak for themselves" (Kitchin, 2014, p. 6). In this research, nearly a million OSI items are collected and analysed using various computational techniques (more details are presented in Sections 1.5.3 and 1.5.4), with an intention to present empirical evidence from massive field data with both data-driven and knowledge-driven methods.

Methodological choice. In general, this thesis adopts a mixed methods approach, but it is quantitative in nature. Mixed methods research integrates quantitative and qualitative approaches by collecting, analysing and mixing both quantitative and qualitative data in a single research project (Bryman & Bell, 2015; Creswell, 2013; Sekaran & Bougie, 2016). According to Saunders, Lewis, and Thornhill (2009), this thesis, in particular, is a mixed-model research, combining quantitative and qualitative techniques and procedures for data collection and analysis. Data collected for this thesis involves both numeric and unstructured textual data, and the qualitative textual data is quantified and converted into quantitative instruments for further statistical analysis (see Section 1.5.4 for details).

Time horizon. This thesis presents both cross-sectional and longitudinal studies. A cross-sectional study investigates a problem at a particular time to describe the situation or to compare across different entities, while a longitudinal study is capable of examining changes and development over time (Saunders et al., 2016). The current research answers the research questions by investigating firm-level phenomena as well as changes and/or impacts along the timeline.

1.5.2 Research design

A research design is an overall plan of action taking into consideration the research's aims, objectives and questions, as well as the constraints of the project (Sekaran & Bougie, 2016). This subsection first illustrates the framework that reflects the dimensions of the research process (see Figure 1-3). Then, an overview of the three studies in the thesis is presented by summarising and comparing topics, the literature base, methods, and implications (see Table 1-2).

As illustrated in Figure 1-2, the model depicts the proposed structure and process flow of the research. The project starts with identifying the big data phenomenon and the need for applying big data analytics in business and management research to achieve competitive advantages. A comprehensive literature review of big data, big data analytics, and big data research in management is performed to establish the gaps in the current knowledge base (see Chapter 2). Subsequently, the needs and scope of this research are defined, which is the area of firms' engagement in OSIs (see Section 1.2 in Chapter 1). To achieve the research aim of filling this gap in the literature, the current research is thoroughly designed for pinpointing firms' positions in the OSI network, examining potential business impacts of firm engagement, and exploring effective engagement strategies to enhance business performance (see Section 1.5 in Chapter 1). The analysis of firm engagement consists of three independent studies outlined below. The research concludes by discussing the theoretical and practical implications and offering directions for future research.

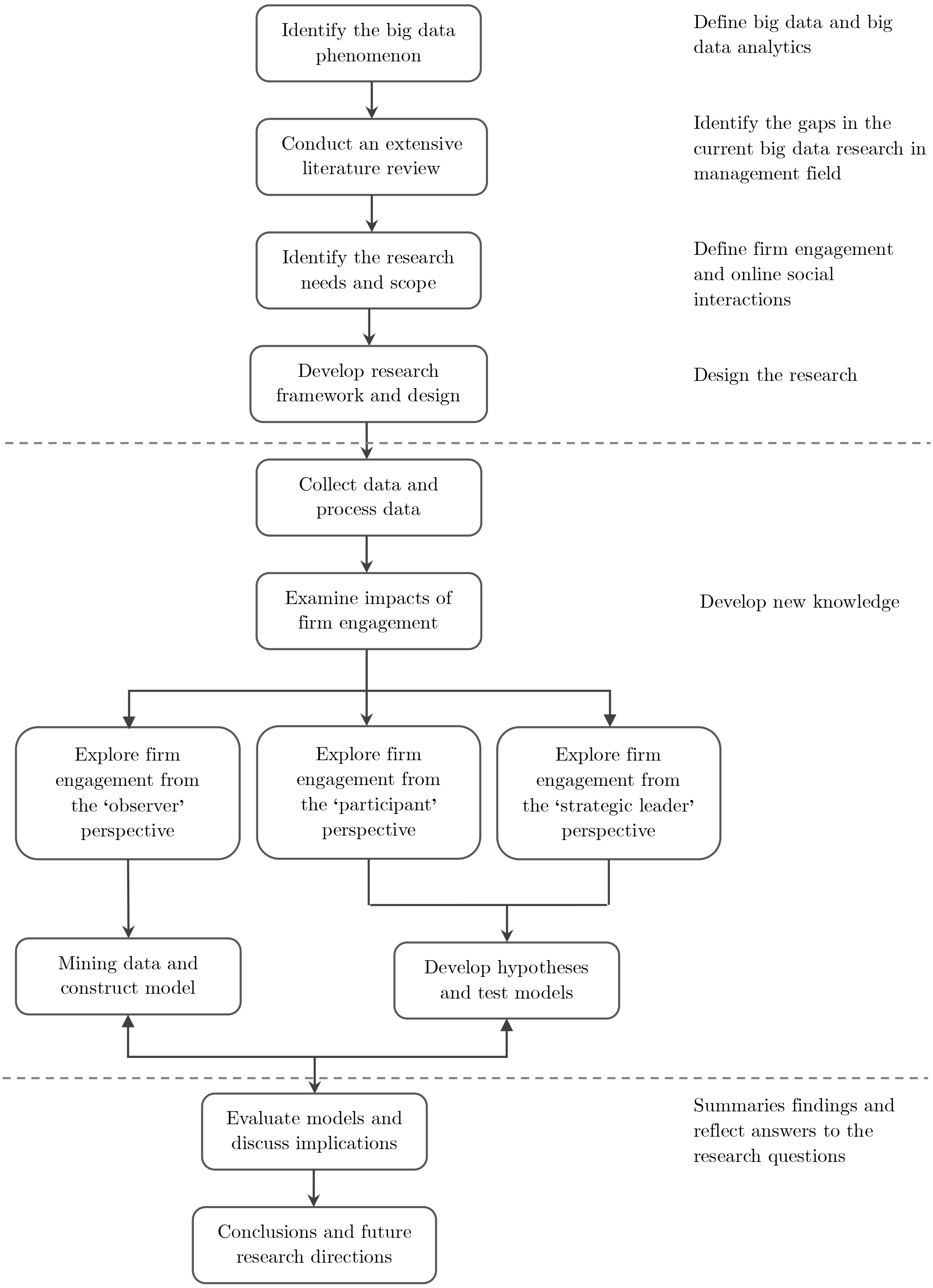


Figure 1-2. Research design

The central theme of this thesis is business engagement strategies in the OSI network. Specifically, this research focuses on the service industry (hotel sector) and online customer reviews (a type of eWOM in OSIs) and empirically investigates how firms act on customer reviews and evaluates the effectiveness of these actions in driving strategic advantages using a range of computational and statistical methods. Corresponding to the three firm engagement roles demonstrated in the research framework, the thesis presents three studies (see an overview in Table 1-2). Each of the three studies independently investigates one firm engagement role by examining potential impacts and developing effective strategies for firms' exploiting online voices.

Table 1-2. An overview of three studies in the thesis

		Study I	**Study II**	**Study III**
	Firm engagement role	Observer	Participant	Strategic leader
	Firm engagement impact	Understand customer needs and opinions	Improve online customer review rating	Encourage customer engagement behaviour
	Strategic value	Service improvement	Customer satisfaction	Customer engagement
	Literature base	Customer co-creation SERVQUAL	Signalling theory Consumer inference	Social influence Social media marketing
	Implications	Discover customer concerns about service quality through mining user-generated content	Develop an action frame (communication styles) for responding to online reviews to as to improve ratings	Establish social media presence and enhance activeness to stimulate customer engagement behaviour
Research methodology	Data source	Online customer reviews and managerial responses		
Research methodology	Data analysis	Topic modelling Machine learning	Text analysis Sentiment analysis Multilevel regression	Statistical analysis Multilevel regression
Research methodology	Approach	Inductive	Deductive	
Research methodology	Methods	Mixed method		Quantitative
Research methodology	Purpose	Exploratory	Explanatory	

Study I explores firm engagement in OSIs from the 'observer' perspective, whereby firms observe and learn from customers' eWOM. This study aims to understand consumers' needs through mining customers' online comments and to demonstrate how to translate consumer experience and opinions towards service quality into a knowledge base for service improvement. Study II explores firm engagement in OSIs from the 'participant' perspective, whereby firms directly join in with online communications. This study aims to examine how firms, in responding to customer reviews, may influence future review ratings and develop effective communication strategies for online reputation enhancement. Study III explores firm engagement in OSIs from the 'strategic leader' perspective, whereby a firm initiates and fosters communication in the virtual community by acting online. This study aims to investigate behavioural effects of firms being present and active online and to suggest the strategic value of business social media activeness in stimulating customer engagement behaviour. Findings from the three studies respond to the research questions about the business impact and its magnitude of firm engagement by taking each of the roles and about the data-driven strategies for effective management of OSIs.

1.5.3 Data collection

Given the particular importance of firm–customer interactions in the service industry, this research uses secondary data from the hotel sector that is publicly available on the Internet. Specifically, the data on online customer reviews and managerial responses of London hotels is retrieved from a leading travel website. The reason for selecting London is that it is a globally popular travel destination. The customer reviews and management responses on a range of hotels in the city meet expectations for sampling and the requirements for research purposes. Although concerns may be raised about potential regional bias by restricting data to a particular geographic area, this sample represents a global leading travel market. The potential biases resulting from random sampling of hotels on the website and variation in sites across different countries are reduced. Besides, the approach of focusing on a specific market is valid because firms in the same region are localised and involved in regional market competition. Studying customer reviews and firms' management styles in a particular region will provide insights for companies engaged in that market to gain competitive advantages.

The characteristics and the entire review and response history of London hotels listed on the site at the time of data collection are downloaded.[3] As such, a representative sample is ensured by substantially reducing the potential selection bias. The collected hotel information includes hotel identity, overall customer ratings, star ratings, and the number of rooms. For customer reviews and managerial responses, all available information is collected, including numerical ratings, review titles and text, reviewer identities, review date, response date, response text, and responder identity. Ultimately, the raw data set contains 813,287 customer reviews and 368,758 attached management responses of 1,063 London hotels over a period of approximately 15 years (December 2001 – March 2016).

The data set is large in volume and contains both structured (e.g., numeric ratings, review and response dates) and unstructured data (i.e., online customer reviews and managerial responses are written documentary data). Its big data nature requires proper data acquisition and efficient management with advanced algorithms and databases. All the collected data is transmitted to, stored and managed in a MySQL database. MySQL is an open-source relational database management system (RDBMS) which has gained global popularity. Such a data centre is cost-effective and capable of handling huge data sets, where data can be easily accessed and updated by running SQL (structured query language) queries. In this research, the database is used to store and pre-process data, including data integration, cleansing and redundancy elimination.

1.5.4 Data analysis

Given the sheer volume of data and the unstructured nature of reviews and responses, various big data analytics techniques are employed in the three studies to analyse the large-scale data. A summary of methods is presented in Figure 1-3.

[3] At the time of data collection (March 2016), there were 1,063 hotels in London. The sample is restricted to hotels with formal accommodation and full services, and excludes other types of accommodation, such as special lodgings, holiday rentals, B&Bs and inns, so as to maintain data consistency and eliminate possible impacts of firm types. The sample ranges from the very first review of each hotel to the latest one at the end of data collection period.

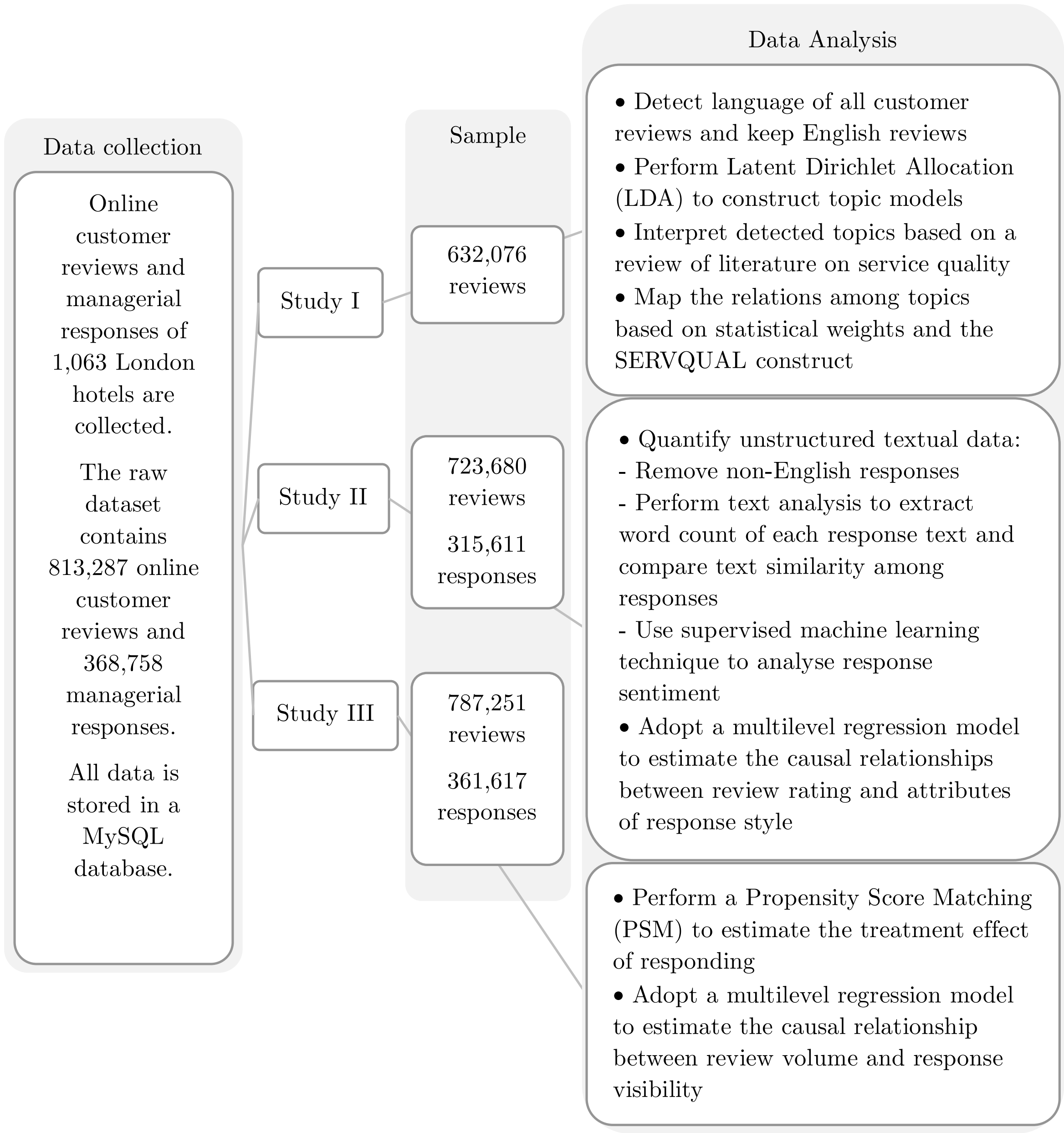


Figure 1-3. A Summary of data collection and analysis methods[4]

Study I adopts an exploratory, inductive and mixed methods approach. The main method used in this study is topic modelling, which is a text-mining technique combining machine learning and natural language processing for detecting underlying thematic information from large archives of documents and organising the unstructured text based on the identified topics (Blei, 2012). This approach helps uncover the hidden semantic structure and extracts topics from the

[4] Sample size in this Figure shows the amount of processed data (e.g., remove non-English reviews, remove duplicates) used in each study.

collection of review documents. There are various topic models, and a Latent Dirichlet Allocation (LDA) model is applied in this study as it is the simplest yet most efficient (Blei & Lafferty, 2009). It detects topics by identifying lists of co-occurring words and the relative importance of topics based on the observable words and their probability distributions. The detected topics are labelled by interpreting lists of similar keywords, and the themes of these topics are validated by comparing against prior studies of determinants of service quality and customer satisfaction. Further, the topics are grouped using the SERVQUAL construct to map the dimesnions of service quality and their relative importance based on the statistical weights of topics. In addition, these quality dimensions are measured in different periods and compared across different market segments and customer groups to demonstrate shifts in consumer perception of service quality. More details about the topic-modelling approach will be presented in Chapter 3, in which it is applied to extract key topics from the sampled customer reviews.

Study II adopts an explanatory, deductive and mixed methods approach. Research methods used in this study include text analysis, sentiment analysis, machine learning, and multilevel regression. The first step is to analyse the textual features of management responses. Specifically, text analysis is performed to extract word counts of each response by tokenising each response document; to compare similarity in response content of a particular hotel, a vector space model is used to generate a cosine similarity matrix which measures how related the responses documents are based on the term frequency–inverse document frequency (tf–idf) schema; sentiment analysis is conducted to identify sentimental orientation in managerial responses through training and applying a linear support vector machine (SVM) learner to the collection of responses and assigning each response a sentiment score. These processes quantify the unstructured response text into quantitative measures. Then, a multilevel regression model is adopted to estimate the causal relationships between review ratings and the quantified response attributes. The panel approach investigates response effects in both cross-sectional and longitudinal dimensions.

Study III adopts an explanatory, deductive and quantitative approach. Statistical analysis and multilevel regression are employed in this study. The first analysis tests the effect of response provision on review volume. A propensity score matching (PSM) technique is adopted to create a treatment group and a comparable control group to estimate the treatment and cross-section effects of responding on the daily number of reviews. Specifically, PSM in this study uses a binominal logistic regression specification to estimate the propensity score and a one-to-one matching in the nearest neighbour algorithm based on the calculated propensity scores to create the treatment and control groups. The second analysis focuses on the responding hotels and adopts a multilevel regression model to examine the causal relationships between review volume and response visibility. It measures the after-treatment ongoing effects of response ratio and response speed across hotel–month panels.

1.6 Research Contributions

This research first contributes to the big data research in the business and management domain. There is a growing awareness of big data's business values and managerial changes led by the data-driven approach, yet the current research frontier of big data in management field lacks clarity (Chen et al., 2012; Sivarajah, Kamal, Irani, & Weerakkody, 2017). The study fills the gap with an interdisciplinary review of big data-related studies, which depicts a clearer path towards management progress in big data value achievement through recognising past accomplishments and identifying fruitful research directions. In so doing, the thesis enriches our knowledge about the potentials of strategising big data. Moreover, instead of using traditional approaches such as experiment or case study, this research applies big data analytics to a large set of quantitative and qualitative data. The mixed methods design and a range of advanced techniques convert massive OSI information into measurable instruments, which enables a more comprehensive analysis of the research problems. The data-driven approach using publically available data captures specifically detailed aspects of customer and business activities, which offers a concrete interpretation of the real-world story (Kitchin, 2014). Therefore, this thesis contributes to management research by applying and generalising the big data approach in the business sphere.

This research also contributes to the strategy and marketing literature. The review of big data research reveals that social media attracts great attention of scholars; nonetheless, few studies have explored how firms should act on the massive online interactions among users. Past studies predominantly examine the OSI issue from the consumer side, with efforts devoted to interpreting how and why customers share information and interact online, and how this influences business performance (e.g., see De Matos & Rossi, 2008). In contrast, only limited studies have examined the actions and influence of firms' involvement and engagement in the

OSI network. This thesis investigates OSIs from a business perspective by identifying different levels of firm engagement and examining consequences of various action plans. This extends prior studies on OSIs (e.g., Godes et al., 2005) and contributes to the social media strategy literature (e.g., Kaplan & Haenlein, 2010) by providing a more comprehensive view of OSIs and empirical evidence on the efficacy of firm engagement strategies. Furthermore, analysis of firm engagement roles and strategies offers practical insights into strategically managing voices from the online customer crowd in the marketplace. The empirical evidence derived from massive field data is highlighted for implementing effective marketing strategies for enhancing customer experience, satisfaction and engagement in the digital business environment.

1.7 Outline of the Thesis

The thesis is organised as follows.

Chapter 1 begins with an introduction of the research background and motivation for this research. This is followed by establishing research questions and objectives and the research framework. Next, the research methodology is described, clarifying the type of research, research design and research methods for data collection and data analysis. Finally, the research contributions and the organisation of the thesis are outlined.

Chapter 2 presents a comprehensive review of the literature on big data and big data analytics in the business and management domain. A survey of big data research over the past decade identifies the key themes in the current body of knowledge and clarifies the research frontiers. The review identifies that firm engagement in OSIs is a fruitful area to study, and this is the knowledge gap that this research aims to address.

The thesis proceeds with three stand-alone studies (Chapter 3, 4 and 5)—each study is presented with an introduction, literature review, methodologies, findings and conclusions. The three studies demonstrate how firms can benefit from engaging with social media through playing three different roles—namely observer, participant and strategic leader in the OSI network respectively.

Chapter 3 presents Study I entitled "A *data analytic approach to transform user-generated content into service improvement*". This chapter explores how firms can harness customers' online reviews to improve service quality. It demonstrates a data-driven strategy for identifying and prioritising dimensions of consumer-perceived service quality by applying an LDA topic model to mine opinions from online conversations. Findings from this study reveal the predominant attributes that affect consumers' perception of hotel services, which highlights the potential for service firms to mine knowledge from UGC to design and improve services.

Chapter 4 presents Study II entitled "*Strategies for responding to customer reviews: Estimating the interactive effects on future ratings*". Using a large-scale data set, this chapter investigates how managerial responses to online reviews influence customer ratings in the future. Combining text analysis and a multilevel regression approach, this study examines the detailed nature of managerial responses and their interactive effects on customer ratings over time. The empirical test presents field evidence of the efficacy of managerial responses and suggests effective communication styles in improving online customer ratings.

Chapter 5 presents Study III entitled "*The strategic value of business online presence and activeness: Stimulating customer engagement in online communications*". This chapter analyses the behavioural effects of firms' online presence and activeness in influencing customer engagement in word-of-mouth communications. It examines the extent to which the provision of managerial responses affects the volume of customer reviews and how such impact varies with the visibility of managerial responses. Findings from this study underline the importance of firm engagement and online activeness in the online communication network in stimulating customer engagement behaviour.

Chapter 6 concludes the thesis. It first presents a summary of findings from the three studies. This is followed by a discussion on the theoretical contributions to big data research in management and digital marketing research, as well as practical implications for business engagement strategies in the OSI network. The last section considers the research limitations and provides suggestions for future research.

CHAPTER 2

LITERATURE REVIEW

2.1 Introduction

This chapter presents a review of big data research across a range of management disciplines. The purpose of the literature review is to scope out the current state of knowledge regarding how big data and analytics has been studied and used in the management field, instead of discussing the theoretical basis for the three empirical studies. The relevant theories (e.g., signalling theory, customer inference theory, social influence network theory) that underpin the individual empirical studies will be discussed in the relevant literature review sections in later chapters.

The literature review starts by defining the concept of big data (Section 2.2). Next, Section 2.3 surveys the literature on big data analytics and summarises the available analytics to deal with big data, which inform this research about advanced techniques and analytics. Then, Section 2.4 reviews management research that surrounds the big data topic or applications of big data analytics, with the aim of synthesising big data studies from various perspectives in management and highlighting the key research interest in each stream. Finally, Section 2.5 illustrates the necessity of strategising big data in business by demonstrating the debates on big data and the associated challenges and opportunities of big data in creating business benefits. This section also discusses the identified research gap of firm engagement in OSIs and the importance of understanding this from a dynamic capabilities view.

2.2 Defining Big Data

> *"Big data refers to things one can do at a large scale that cannot be done at a smaller one, to extract new insights or create new forms of value, in ways that change markets, organisations, the relationship between citizens and governments, and more."*
>
> — Mayer-Schönberger & Cukier, 2013, p. 6

Big data, as evident in the term, implies an enormous amount of data. IBM states that big data is significantly large-scale and complex to handle with traditional databases and tools. But there seems to be no fixed threshold for the measurement of what size or type of data can be treated as big data. This is due to the amount and diversity of data, which continues to exponentially increase as the Internet of Things matures. In a forward-looking sense, some scholars suggest that big data is a "moving definition" that varies with time and industrial sectors (Manyika et al., 2011, p. 1). Although the concept has not been rigorously defined, big data is not only about size. An emerging consensus about the uniqueness of big data, which distinguishes itself from what we recognise as large data sets, appears to have been achieved across researchers and practitioners.

Big data has a few distinct features. At the beginning of the development of the big data concept, Laney (2001) introduces the three Vs of big data, namely *volume*, *variety* and *velocity*, in order to illustrate the continuous expansion of data in terms of multiplicity (see also Kwon, Lee, & Shin, 2014; Russom, 2011). First of all, the data volume is exploding with information created and stored in digital formats. In addition to this scale expansion, data variety is also increasing. Data is generated and collected from diversified sources such as web sites, smart devices and social media. The format of data is no longer limited to structured data that can be neatly mapped into relationship databases. Unstructured and semi-structure data is also identified as useful and usable forms of data that contain abundant information. Furthermore, the speed of data generation and data delivery is faster than ever before. Torrents of data are generated in batches, real-time or in streams, which require data processing and analysis to accelerate in tandem so as to act on the high-frequency and perishable data. Additional dimensions of big data, including high *variability*, low *veracity* and high *value*, further enrich its conceptualisation (see Gandomi & Haider, 2015; Jin et al., 2015; Katal et al., 2013). These characteristics emphasise the inconsistency and uncertainty of big data and hence advanced techniques are urgently needed to ensure the reliability and validity of insights derived from big data to make business impacts.

The 6Vs nature of big data pinpoints the complexity that businesses face in regard to creating real value through mining insightful information from copious data. In this doctoral research, the term 'big data' refers to an extremely large amount of data in disparate formats continuously generated from multiple sources. The data can either be large structured data sets or unstructured/semi-structured data, such as text data (e.g., documents, natural language), web data (e.g., web structure, web usage, web content), social media data (e.g., virtual communities), multimedia data (e.g., image, audio, video), or mobile data (e.g., sensor, geographical location, mobile application) that is generated by machine, human behaviour and interactions.

2.3 Big Data Analytics

2.3.1 The rise of big data analytics

The abundance and diversity of big data have been regarded as information assets that can yield insights for businesses to act on (Russom, 2011). But most organisations have massive data without sufficient understanding and capabilities to use it effectively (LaValle et al., 2011). In recent years, the proliferation of data has inspired practitioners and academics to take advantage of it with more effective analytics (Phillips, 2017). In this regard, big data analytics has emerged, which involves applying advanced tools and analytics techniques to manage and analyse complicated big data sets (LaValle et al., 2011). Compared to traditional business intelligence, big data analytics has an improved capability to handle big data and discover new business facts from massive data (Russom, 2012). It has become a trendy practice in business analytics (Russom, 2011).

Technology improvement is critical to big data analytics in terms of inventing advanced hardware and software, platforms and tools to collect, store and process big data, which enhances efficiency and cost-effectiveness in data management (Turban et al., 2008; Watson & Wixom, 2007). In addition, big data analytics relies on advances in analytics techniques. Advanced analytics moves beyond a rear view of describing past events towards diagnostic, predictive and prescriptive analysis to drive business decisions and future actions (Delen & Demirkan, 2013). By employing data mining, simulation, modelling and many other advanced analytics techniques, big data analytics presents better accuracy in terms of predictions based on solid data evidence instead of intuition and experience (McAfee & Brynjolfsson, 2012).

Given the importance of technology support and advanced analytics techniques in bridging big data and business decisions, this section reviews related scholarly work on big data analytics to identify technological advancements and emerging analytics. The following review takes both a technical and a managerial perspective. The technical view follows the idea of the big data value chain in Chen, Mao, and Liu (2014) to demonstrate key technologies that address big data challenges. The managerial view adopts the taxonomy of emerging analytics areas introduced in Chen et al. (2012) to explore the current research fronts in regard to the application of big data analytics in the business management community.

2.3.2 Advanced technology

The review first focuses on the technical aspect of big data analytics. Given that the nature of big data presents technical challenges to data management and data analysis, efficiency is a primary goal in handling complex data sets. Improved technological infrastructure and platforms improve the capability to manage data at every stage of the analytics workflow. Indeed, new techniques are constantly being proposed and discussed in fields such as computer science and engineering. These techniques can help achieve better data quality, adequate storage space, faster access and processing speeds, deeper analysis and powerful results presentation. By applying these cost-effective and efficient technologies, big data becomes manageable and valuable information that can potentially be extracted to inform business decisions. The following discussions introduce the advanced technologies applied in big data analytics based on the analytics workflow structure, around the theme of analytics efficiency.

2.3.2.1 Big data acquisition

Data sources have become diversified with massive data being generated from machines, devices, human interactions and countless other sources (Hu, Wen, Chua, & Li, 2014). Data acquisition is concerned with not only the collection of raw data from data sources, but also the transmission and pre-processing of data (Chen et al., 2014; Hu et al., 2014). Khan et al. (2014) indicate that data collection, filtering and classification are the main tasks at this stage. Given the large volume and low veracity of data, targeting useful data and cleaning it to reduce data inconsistency and improve data quality are particularly vital for carrying out meaningful analytics.

To retrieve raw data generated from various sources, new techniques have been developed. Data can be automatically collected from sensors (e.g., Jonsson & Eklundh, 2002; Luo, Wu, Sun, & Chen, 2009; Wang & Liu, 2011), mobile devices (e.g., Baak, Müller, Bharaj, Seidel, & Theobalt, 2013; Kaplan & Hegarty, 2005; Laurila et al., 2012), and radio frequency identification (RFID) (e.g., Roberts, 2006). We can use distributed architecture to capture system log files (e.g., Moreta & Telea, 2007; Nicholas & Huntington, 2003; Suneetha & Krishnamoorthi, 2009; Thelwall, 2001b) and web crawlers have been developed (e.g., Choudhary, Dincturk, Mirtaheri, Moosavi, Von Bochmann, Jourdan, & Onut, 2012; Kaplan & Haenlein, 2010; Thelwall, 2001a) for mining unstructured documents and network data (e.g., Liu, Yang, & Zhang, 2013) from web sites.

Data transmission refers to the transportation of data into a data storage infrastructure or data centre. It can be Internet backbone transmission and data centre transmission. In particular, the inter data center network (DCN) (e.g., Ghani, Dixit, & Wang, 2000; Shieh, 2011) and intra-DCN (e.g., Barroso, Clidaras, & Hölzle, 2013) allow mobility of the data within or between storage devices. A final and key step to improve data quality is data pre-processing, which requires data integration (e.g., Gravano, Ipeirotis, Koudas, & Srivastava, 2003; Lenzerini, 2002), cleansing (e.g., Jeffery, Alonso, Franklin, Hong, & Widom, 2006; Maletic & Marcus, 2000) and redundancy elimination (e.g., Hussain, Asghar, & Masood, 2010; Sarawagi & Bhamidipaty, 2002; Tsai & Lin, 2012). At this stage the data is aggregated into a uniform format and the processed data has benefits for big data analytics due to their increased consistency and reliability (Chaudhuri, Dayal, & Narasayya, 2011).

2.3.2.2 Big data storage

One of the key observations is that data volume is experiencing spectacular growth, generating a need for increased data storage of a higher standard in terms of space and ease of access. Advanced data storage techniques can maximise storage space using a distributed or networked infrastructure and help to save costs and time by conducting analysis within the database or memory.

Data storage capabilities rely on hardware infrastructure, database and management, and programming models (Hu et al., 2014). To store larger data sets, storage architectures have been advanced with enhanced capabilities, such as network attached storage (NAS), storage area network (SAN), and direct attached storage (DAS) (e.g., Barker & Massiglia, 2002; Chen et al., 2014; Gibson & van Meter, 2000; Khan et al., 2014; Reed, Chron, Burns, & Long, 2000; Shiroishi et al., 2009; Telikepalli, Drwiega, & Yan, 2004). These storage facilities connect to each other or link to a network, which gives easier access to other storage space. Moreover, distributed storage frameworks, such as Google File System and Hadoop Distributed File System, have advantages in dealing with large data sets and streaming data with cheaper disk drivers (e.g., Ghemawat, Gobioff, & Leung, 2003; Shvachko, Kuang, Radia, & Chansler, 2010). Data is broken down into smaller scales and distributed in different servers, which enables scalability for processing.

As for the databases used to store and manage data, relational database management system (RDBMS) is mainly used for large-scale structured data sets (Moniruzzaman & Hossain, 2013; Ramakrishnan & Gehrke, 2000). For non-relational databases, there are three major types, namely key-value databases, document databases, and column-oriented databases. It is faster to retrieve data with key-value pairs, and easier to compress data and operate parallel processing with data records in a sequence of columns (e.g., Cattell, 2011; Chodorow, 2013; DeCandia et al., 2007; McCreary & Kelly, 2013). In-memory (e.g., Hahn & Packowski, 2015; Watson, 2014) and in-database (e.g., Russom, 2012; Chaudhuri et al., 2011) techniques enable data processing and analysis within the memory or database instead of transporting data between the data centre and disks. In addition, public or private cloud computing infrastructures also facilitate data management. Resources in clouds can be allocated dynamically, but this causes security concerns (e.g., Assunção, Calheiros, Bianchi, Netto, & Buyya, 2015; Bi & Cochran, 2014; Talia, 2013).

2.3.2.3 Big data processing

According to Elgendy and Elragal (2014), there are several requirements in processing big data, including fast speed in terms of data loading query processing, high efficiency in storage space utilisation, and strong adaptability to the dynamic workload. Timely processing techniques can speed up data processing and enhance the efficiency of large-scale data analytics. Researchers have begun to examine a range of processing techniques such as generic processing models (e.g., Condie, Conway, Alvaro, Hellerstein, Elmeleegy, & Sears, 2010; Dean & Ghemawat, 2008, 2010; Grolinger, Hayes, Higashino, L'Heureux, Allison, & Capretz, 2014; Sagiroglu & Sinanc, 2013), stream processing models (e.g., Cherniack, Balakrishnan, Balazinska, Carney, Cetintemel, Xing, & Zdonik, 2003; Neumeyer, Robbins, Nair, & Kesari, 2010; Stonebraker, Çetintemel, & Zdonik, 2005), and graph processing models (e.g., Lumsdaine, Gregor, Hendrickson, & Berry, 2007;

Malewicz, Austern, Bik, Dehnert, Horn, Leiser, & Czajkowski, 2010; Salihoglu & Widom, 2013). The latter two can deal with large-scale data using graph and event nodes.

In terms of queries, compared to relational processing, which uses a structured query language (SQL) to access structured data in a relational database (e.g., Pedersen & Jensen, 2001), parallel processing can perform more efficient queries. Data is spread across a large number of servers and computing problems are solved on separate servers in parallel. No memory or resources need to be shared across different servers and the system can be expanded with additional servers. This is highly efficient for processing large-scale data sets and unstructured data (e.g., Jordan & Alaghband, 2002; Parhami, 2006; Roosta, 2012).

2.3.2.4 Big data analysis

Data analysis seeks to discover facts, understand relationships, predict the future and advise on possible outcomes (Khan et al., 2014). Accordingly, data analysis can be descriptive, predictive and prescriptive (Hu et al., 2014). A wide range of analysis techniques have been adopted in analysing large data clusters. These include well-developed methods such as statistical analysis and regression to discover causality and effect as well as emerging approaches such as data mining and simulation to predict future outcomes (Delen & Demirkan, 2013).

A review of the research in the recent decade reveals that advanced methods are increasingly being adopted, including data mining, text mining, web mining, machine learning, simulation and network analysis, among others (see Han, Kamber, & Pei, 2011; Singh, 2014; Witten, Frank, & Hall, 2011; Wu et al., 2008). These analysis methods rely on computational algorithms and expert knowledge to perform an in-depth examination. In addition, to present a massive amount of data and an analysis of the results, concise and interactive platforms are required, such as advanced data visualisation. These platforms are helpful in assisting data users to better engage with data analysis and interpretation.

2.3.3 Advanced analytics

Following the above discussions on the advances in data analysis approaches, this subsection investigates the application of advanced analytics in business and management disciplines. According to Chen et al. (2012), in the business intelligence 1.0 system, the data is mostly structured with pre-defined indexes or primary keys indicating the relations among data entities. The review indicates that techniques for structured data analytics have experienced tremendous improvements in adapting to big data and new technologies. Novel platforms and advanced analytics have been developed, while a number of mature programmes in current business intelligence systems continue to serve in big data analytics by continuously optimising the designs for handling large volumes of data.

The review also reveals that current research emphasises unstructured data (Chen et al., 2014; Hu et al., 2014). In the business intelligence 2.0 and 3.0 systems, data is generated with a faster speed and greater diversity, which requires new analytics techniques to meet new challenges in analysing structured and unstructured data. The following discussions on business-related studies underline the growing interest in unstructured data analytics. Following the classification approach in Chen et al. (2012), the related literature is grouped based on various types of unstructured data, namely text, web and multimedia data, network data and mobile data.

2.3.3.1 Text analytics

The first stream of research is around text analytics. Text analytics depends largely on text mining, which deals with unstructured text from documents, emails, logs, web pages, social media, comments, feedback and so on. For text analytics, document representation and query processing help retrieve information and natural language processing (NLP) detects certain words, phrases, events and topics from massive data. Then the processed data can be used to construct models for further analysis, such as topic models (e.g., Blei, 2012), language models (e.g., Kao & Poteet, 2007), and sentiment or opinion mining (e.g., Garg & Chatterjee, 2014; Pang & Lee, 2008). Hence, information retrieval and statistical NLP are the basis upon which more text-based searching and analysing techniques are innovated. In reviewing the relevant literature, it is found that a great number of studies focus on information retrieval (e.g., Beebe, Clark, Dietrich, Ko, & Ko, 2011; Duan, Zhang, Wang, & Xu, 2011), clustering (e.g., Aliguliyev, 2009a; Wei, Chiang, & Wu, 2006), classification (e.g., Chou, Sinha, & Zhao, 2010; Colace, De Santo, Greco, & Napoletano, 2014), and summarisation (e.g., Aliguliyev, 2009b; Baralis, Cagliero, Jabeen, Fiori, & Shah, 2013) of textual data. These text mining methods are widely applied to detect opinions (e.g., Mostafa, 2013; Ordenes, Theodoulidis, Burton, Gruber, & Zaki, 2014), idea mining (e.g., Thorleuchter, Van den Poel, & Prinzie, 2010), personalisation (e.g., Fan, Gordon, & Pathak, 2006; Wei, Yang, & Hsiao, 2008), recommendation (e.g., Hyung, Lee, & Lee, 2014; Zhang & Jiao, 2007), market analysis (e.g., Nassirtoussi, Aghabozorgi, Wah, & Ngo, 2015; Netzer, Feldman, Goldenberg, & Fresko, 2012), fraud detection (e.g., Glancy & Yadav, 2011), and many others.

2.3.3.2 Web and multimedia analytics

Web analytics has become an essential part of Web 2.0 systems (O'Reilly, 2007), which aim to discover and analyse useful information from web documents and services, such as web text, link structures, web logs and other types of web data (Ashton, Evangelopoulos, & Prybutok, 2014). The web content, structure and usage data (Pal, Talwar, & Mitra, 2002) provide abundant information about user dialogues and behaviour. Mining this information is useful for improving operation and marketing efficiency, such as knowledge discovery (e.g., Chung, Chen, & Nunamaker, 2005), real-time awareness (e.g., Castellanos, Gupta, Wang, Dayal, & Durazo, 2012), personalisation (e.g., Ho, Bodoff, & Tam, 2011), customer prediction (e.g., Yeh, Lien, Ting, & Liu, 2009), inventory management (e.g., Huang & Van Mieghem, 2014), and online advertising (e.g., Chatterjee, Hoffman, & Novak, 2003).

In addition to web data presented in text format, other information contained in images, audio, and video has become the new focus of research. Research on multimedia analytics mainly concentrates on summarisation (e.g., Ding et al., 2012), annotation (e.g., Wang, Ni, Hua, & Chua, 2012), index and retrieval (e.g., Lew, Sebe, Djeraba, & Jain, 2006) and recommendation (e.g., Park & Chang, 2009). These techniques help extract hidden information from images, audios and videos, which can be used with other web and media data to suggest particular content to users based on their preferences. Compared to other types of data, only a handful of studies use multimedia data to gain business insights. Part of the problem is the richness of the content, which brings difficulties in extracting information and organising it into structured a relational structure.

2.3.3.3 Network analytics

Network science is evolving with the rapid growth in online interactions and online social networking. The massive amount of user-generated data often reflects consumers' opinions and connections among entities and communities. The review illustrates that social media analytics and network analytics seem to be a promising direction in big data research. It covers the analysis of social media activities, user behaviour and relationships, and semantic and textual information in user-generated content (UGC). A number of new techniques are being applied for network analytics. For example, researchers are focusing on sentiment analysis to detect sentimental orientation and opinions conveyed by UGC. Different approaches have been proposed for emotion detection (e.g., Balahur, Hermida, & Montoyo, 2012) and sentiment classification (e.g., Colace, Casaburi, De Santo, & Greco, 2015; Da Silva, Hruschka, & Hruschka Jr, 2014), which are applied to understand customer satisfaction and opinions (e.g., Alfaro, Cano-Montero, Gómez, Moguerza, & Ortega, 2016; Kang & Park, 2014). Besides, social network analysis is often performed to map relations in the online community. Network analysis discovers potential links, social influences and interactions in the network, which is helpful in predicting consumer behaviour (e.g., Cascio et al., 2015; Lee, Hosanagar, & Tan, 2015), detecting social influence in the virtual community (e.g., Lu, Jerath, & Singh, 2013; Sridhar & Srinivasan, 2012), and enhancing brand and sales performance (e.g., Goh, Heng, & Lin, 2013; Wang, Ting, & Wu, 2013).

2.3.3.4 Mobile analytics

Mobile devices are penetrating universally and mobile computing technologies enable millions of applications to generate a vast amount of information. Mobile and sensor-based systems foresee a promising future in business intelligence and analytics. Relying on applications embedded in smart devices such as smartphones and tablets, data generated from these sources are usually fine-grained, location-specific, context-aware and highly personalised. The review shows that, typically, mobile analytics uses mobile sensing, web logs, networking and applications to acquire personalised data from mobile devices and create behavioural models of individuals in order to realise targeting (e.g., Fong, Fang, & Luo, 2015; Luo, Andrews, Fang, & Phang, 2014), advertising (e.g., Andrews, Luo, Fang, & Ghose, 2016), personalisation (e.g., Chung, Wedel, & Rust, 2016; Lee, 2007) and recommendations (e.g., Yang, Cheng, & Dia, 2008). This provides great opportunities to innovate and advance our understanding of markets and customers in a timely manner.

2.3.4 Summary of big data analytics

Within the review scope, the investigation into the collected studies has led to the identification of two dimensions in the current research on big data analytics. From a technical perspective, prior studies focus on the analytics workflow, in which advanced technologies can help improve efficiency in handling big data. From a managerial perspective, advanced techniques target diversified data types and sources and assist business analytics with predictive and prescriptive analytics to derive insights.

One of the key observations from the review in this section is that unstructured data is being generated and collected on an astronomical scale across various industry sectors. Among the four types of emerging analytics of unstructured data, text analytics and network analytics seem to be more attractive to management researchers than web, multimedia and mobile data analysis. This may be a result of the growing number of users on social media platforms (e.g., Facebook and Twitter) and the ease of accessibility of the data sources (Chan, Lacka, Yee, & Lim, 2017; Mount & Martinez, 2014; Rapp, Beitelspacher, Grewal, & Hughes, 2013). The abundance of text and social media data offers promising avenues for future research regarding how to effectively utilise such data by performing meaningful analytics to address business problems and inform decision-making.

The review reveals that perhaps the most prominent gap in the current literature is the limited attention paid by scholars in the management community to the potential of harnessing big data analytics to achieve competitive advantage. Studies in information management often propose novel approaches to handling big data but these studies cannot flourish in isolation. Fulfilling the potential of big data requires the application of these techniques in marketing and operations research as well as inputs from other management disciplines such as strategy, innovation, and organisation. Therefore, research is now needed to advance our understanding and utilisation of big data analytics in regard to managerial applications. This requires a combination of interdisciplinary knowledge and a mixed methodological approach to address the challenges presented by big data when making use of it (Chan, Wang, Lacka, & Zhang, 2016).

2.4 Big Data in Management Research

This section presents a review of the literature in the business and management community in order to identify key themes in the current research, which discusses big data and performs big data analytics in deriving business insights. To survey the existing big data research, the review in this section takes a multidisciplinary perspective to uncover the research streams in different management domains, including information management, organisation studies, operation management, marketing and others. The primary research interest and current state of knowledge regarding big data in each domain are discussed. A special attention has been paid to the area of marketing because this thesis follows several research themes (e.g., eWOM, sentiment analysis, social media engagement) in the big data marketing literature. A comprehensive understanding of the relevant studies in these fields lays a solid foundation for carrying out this research.

2.4.1 Information management perspective of big data

The availability and feasibility of information are critical to organisational success in term of strategic decision-making. Accordingly, from the information management perspective, big data research is primarily concerned with data acquisition and data process effectiveness. Information management professionals and researchers are particularly interested in acquiring information and knowledge from various data sources for business use. Studies in this discipline often propose or apply novel approaches to deal with big data for various information management purposes, such as information retrieval, knowledge discovery, text mining, web mining, topic detection, document clustering and document classification.

First, scholars are seeking to improve the effectiveness of *information retrieval*, especially from unstructured data. For example, Wei, Hu, Tai, Huang, and Yang (2007) find that a topic-based method for query expansion is more effective to address word mismatch problems in information retrieval. Beebe et al. (2011) develop an extended approach with a digital forensic text string search process that reduces information retrieval overheads. Liao, Yang, Li, Wang, Qi and Zhu (2014) propose a framework for image retrieval that can efficiently support content-based similarity and semantic searches in cloud datacentres. Besides, new methods of *web mining* have been proposed to improve efficiency in searching for and extracting information. For example, Wang, Lu, and Zhang (2007) develop a method to effectively extract key information from web pages with noisy information. Chau and Chen (2008) propose a method combing web content and structure analysis to filter web pages and search for information. Yang (2009) proposes a method to automatically generate metadata for web pages in semantic analysis.

Thorleuchter and Van den Poel (2013) apply a novel approach to automatically identify new ideas from the web, which facilitates technological innovation. These advanced information retrieval frameworks can help with web searching (e.g., Chung et al., 2005) and knowledge discovery (e.g., Lee & Wang, 2012; Lee, Yang, & Wang, 2011).

In addition, *text mining* has attracted a great deal of attention from scholars in the information management field. Text mining relies on a number of clustering and classification techniques and novel approaches have been put forward in an attempt to enhance the accuracy of managing and extracting information from textual data (e.g., Colace et al., 2014; Duan et al., 2011; Hashimi, Hafez, & Mathkour, 2015; Kou & Lou, 2012; Singh, Hillmer, & Wang, 2011; Ur-Rahman & Harding, 2012; Wei et al., 2006). The proposed algorithms improve document classification and clustering accuracy, which helps with topic detection for business purposes. For example, Wei, Yang, and Lin (2008) adopt a novel multilingual document clustering method to generate knowledge maps. Lo (2008) develops an auto classification mechanism to identify complaints from customer messages so as to illustrate real-time changes in service quality. Thorleuchter et al. (2010) propose an idea mining approach to automatically discover new ideas from textual information. Yoon (2012) finds that a keyword-based text mining method is useful for identifying weak signal topics for future business planning. Bao and Datta (2014) analyse corporate textual data on risk disclosure using text.

Overall, big data research in information management has explored new ways to deal with massive data and these methods can be applied to enhance the efficiency and accuracy of information retrieval and processing. The proposed algorithms also provide solid technical or technological support for research in other management fields.

2.4.2 Organisation perspective of big data

The organisational studies' perspective on big data examines organisational alignment with the data-driven strategy in every aspect of organisational structure, culture, operation and function. It emphasises the potential changes in the organisational ecosystem and management process under the influence of big data strategy. A few research streams have been identified from this perspective of big data.

The first topic is concerned with the impact of big data on enhancing *business intelligence* and business performance. Big data provides numerous interesting correlations, among which the significant correlations can be used to discover and establish the causality with models (Bollier & Firestone, 2010). For instance, firm performance in a competitive environment is significantly correlated with UGC (Sabnis & Grewal, 2015) and social media metrics significantly indicate firm equity value with stronger and faster predictive relations than traditional online behavioural metrics (Luo, Zhang, & Duan, 2013). Another example is that streaming unstructured web data in near real time improves enterprises' situational awareness and thus their operational business intelligence (Castellanos et al., 2012). Indeed, such business insights are acquired by performing big data analytics, which has become an essential element for organisations to gain success (Akter, Fosso Wamba, Gunasekaran, Dubey, & Childe, 2016; Dutta & Bose, 2015). Insights from big data metrics are more rigorous, which endows firms with advanced business intelligence capability (Chen et al., 2012). This creates business value in knowledge management and develops competitive advantage by enabling more effective decisions with greater flexibility and promptness (e.g., Bhimani, 2015; Fosso Wamba et al., 2015; Côrte-Real, Oliveira, & Ruivo, 2017).

Second, a parallel interest in the subject has emerged among strategy scholars who have explored how big data influences *strategic management*. Hitt, Ireland, and Hoskisson (2011, p. 6) define strategic management as "the full set of commitments, decisions, and actions required for a firm to achieve strategic competitiveness and earn above-average returns". From a strategic management stance**,** big data is increasingly admitted as enterprise assets, which is critical to organisational success (Russom, 2011; Dutta & Bose, 2015). This perspective on big data focuses on how resources and capabilities are assembled and utilised to help firms make quality future and long-term decisions (see Hitt, Ireland, & Hoskisson, 2014; Ireland, Hoskisson, & Hitt, 2012). Big data analytics has a great impact on strategic processes and improves consequent decisions by providing new data, insights and actions. In particular, big data shows a broader view of the information flow that comprehensively reflects the potential changes in business operation in real-time (Bhimani, 2015). Thus, it enables executives to extend their knowledge about the business and thereby enhance the effectiveness and flexibility of strategic decisions in a timely and efficient manner (McAfee & Brynjolfsson, 2012). Moreover, big data can improve innovations in business models through data monetisation and digital transformation (Woerner & Wixon, 2015). For example, a digital business strategy, a fusion between an IT strategy and a business strategy, may provide greater insights by considering the scope, scale, speed and sources of such a strategy (Bharadwaj, El Sawy, Pavlou, & Venkatraman, 2013). More importantly, the profound impact of big data is expected to be seen in an organisational strategic transformation (Provost & Fawcett, 2013) and it may reshape the operational mechanism within the organisation as well as its strategic actions (Beath et al., 2012; Mayer-Schönberger & Cukier, 2013). With such data-driven business models emerging, organisations are required to embrace big data and adopt advanced information systems to enhance the effectiveness and flexibility of their decision-making process (Fosso Wamba et al., 2015).

The third line of research discusses *organisational changes* in the transition to a data-driven culture. Within an organisation context, adopting a big data-enabled decision-making strategy has had a great influence on organisational management (Rifkin, 2014). Such a change is achieved via "transforming processes, altering corporate ecosystems, and facilitating innovation" (Brown, Chui, & Manyika, 2011, p. 26) and it has already generated new managerial perceptions. Undoubtedly, there will be a transition of managers' views on experience, expertise and management practice (McAfee & Brynjolfsson, 2012, p. 4). Kiron and Bean (2013) indicate that data-driven decision-making is a promising trend and that the primary factor attributed to successful big data utilisation is organisational alignment in every aspect of companies. Entrepreneurs and top management are encouraged to seize opportunities by making a commitment and strategic adjustments within their organisations (Braganza, Brooks, Nepelski, Ali, & Moro, 2017; Gunasekaran, Papadopoulos, Dubey, Fosso Wamba, Childe, Hazen, & Akter, 2017). For instance, Ren, Fosso Wamba, Akter, Dubey, and Childe (2016) point out that the information system and quality are critical to enhancing business value and firm performance. Greater transparency and accessibility of data are required and technological innovation has become an essential element to support organisational decisions with sophisticated algorithms (Gupta & George, 2016; Sivarajah et al., 2017; Fosso Wamba et al., 2017).

Lastly, organisational studies have also investigated how big data can improve *human resources* practices, such as performance evaluation, training and development, staff utilisation and skills formation (see Scullion & Linehan, 2005). For example, social media can be used to assist with personnel decisions. Roth, Bobko, Van Iddekinge, and Thatcher (2013) find that in frontline recruitment, more insights about potential hires can be revealed through looking at candidates' social media profiles and online professional networks. But Van Iddekinge, Lanivich, Roth, and

Junco (2016) suggest that social media information regarding job applicants may be invalid and that organisations should use the information with great caution (Kluemper & Rosen, 2009). Besides, workforce data can be adopted to assess employee performance, and can deliver more informed solutions to frontline employee management in regard to staff engagement and productivity enhancement (e.g., Lam, Sleep, Hennig-Thurau, Sridhar, & Saboo, 2017; Shah, Irani, & Sharif, 2017). Moreover, as big data requires advanced data analytics skills, talent management and leadership are expected to function in a different way. Tambe (2014) illustrates that labour with technical skills together with big data investment are necessary attributes for productivity growth and development. This means that strategic workforce planning and leadership need to be more engaged in developing the strategic orientation and data-driven operation.

To sum up, within the organisational studies, attention has been paid mostly to the adaptation of various functions in an institution to the data-driven environment and approach. It is suggested that business leaders should engage in the entire big data ecosystem to earn more benefits from sharing information with external bodies (Gerhardt, Griffin, & Klemann, 2012). To further develop the understanding in this area, more specific action plans are expected to guide the process of organisational alignment with big data in the aspects of culture, management functions and operations.

2.4.3 Operation perspective of big data

From an operation perspective, big data has been explored to improve production and operation efficiency. The current research interest is mainly in the areas of operation management (e.g., Chan et al., 2016; Chan et al., 2017) and supply chain management (e.g., Hazen, Skipper, Boone, & Hill, 2016; Kache & Seuring, 2017).

Operation management deals with the design and control of the overall production and operation process to achieve high efficiency in resource planning and effectiveness in business goal accomplishment. With real-time data collected from smart devices, managers have access to unstructured data sources, which is helpful in achieving better performance through understanding customers, managing risks, and reducing costs. The current research has discussed the possibility of incorporating social media, and web and sensor data to improve the *operations process*. For example, web data such as clickstreams can be used to predict offline orders and reduce inventory costs (Huang & Van Mieghem, 2014). Introduced in Guo, Wong, and Guo (2014), cloud-based order tracking and allocation systems can reduce production costs and improve efficiency. Li, Wang, and Bai (2015) leverage process documents for data extraction and task identification and the proposed approach enhances accuracy and efficiency in process mining. Zhong, Huang, Lan, Dai, Chen, and Zhang (2015) explore the use of RFID data to support logistics planning and scheduling. Aloysius, Hoehle, and Venkatesh (2016) study the mobile checkout process in a retail environment to examine the customer checkout experience and potential benefits for retailers. Chan et al. (2017) find that social media comments can be analysed to discover useful information for operation management applications such as process design.

Besides, prior research has investigated the impact of big data on *production and innovation*. For example, Abrahams, Fan, Wang, Zhang, and Jiao (2015) find UGC in social media can be analysed to extract signal cues so as to inform product defect discovery. Kumar, Shankar, Choudhary, and Thakur (2016) develop a mechanism for fault detection in cloud-based manufacturing, which improves manufacturing quality and reduces the cost of product testing. Furthermore, big data promotes innovation in terms of new ideas and methods to collect, store, analyse, adopt, and share data, and it has benefits for the open innovation process (Mount & Martinez, 2014). Chan et al. (2016) discover the usefulness of social media data in new product development and that the data involved in this process helps firms identify consumers' preferences and needs (Zhan, Tan, Li, & Tse, 2016). Similarly, Jin, Liu, Ji, and Liu (2016) explore new opportunities of big consumer data such as online opinions and behaviour for fulfilling customer requirements and developing market-driven product design. It is expected that leveraging big data in business operations can improve production and operation efficiency. Accordingly, operational research should interact more with big data analytics to take advantage of the comprehensive ecosystem.

In addition, for *supply chain management*, big data analytics has proven its value in managing the flow of goods and services. For example, Chae (2015) analyses tweet data to discover insights for improving supply chain practices such as supply chain risk management and demand shaping. Li and Wang (2017) find that sensor data can be used to improve performance in chilled food supply chains and decisions on dynamic pricing. Indeed, new analytics techniques have been developed to support supply chain operations and innovation (Tan, Zhan, Ji, Ye, &

Chang, 2015). Fuchs and Otto (2015) demonstrate that information technology in supply chain planning has significant value and that improvements in the supply chain can be obtained by financing the IT function. There is potential for growth in regard to in-memory analytics applications in supply chain management given the availability of real-time data and integrated processing models to support operations (Hahn & Packowski, 2015). Correspondingly, Wang, Gunasekaran, Ngai, and Papadopoulos (2016) stress the need for businesses to perform supply chain analytics in the logistics and supply chain management to enable better strategy and operations. But it is worth noticing that data quality is very important in research and practice in regard to supply chain management, which requires interdisciplinary cooperation to develop control methods (Hazen, Boone, Ezell, & Jones-Farmer, 2014).

In general, efficiency is the keyword for the operation perspective of big data. It leads to greater utilisation of real time data to reduce costs and risks and to stimulate innovation. The extent to which efficiency can potentially be enhanced with big data incorporated in the existing operation system could be a possible direction for future studies. In addition, the innovative aspect of big data in developing new ideas and new products or services needs further exploration.

2.4.4 Marketing perspective of big data

Through the marketing lens, understanding consumers is the core element in making marketing decisions. A wide range of topics have been found in the survey of the relevant literature—a table is provided to summarise the main research areas and to list key studies around these topics (see Table 2-1). In general, the marketing perspective of big data focuses on three broad streams, including understanding consumer behaviour, interpreting consumer sentiments and developing data-driven marketing strategies.

2.4.4.1 Consumer behaviour

Consumer behaviour reflects the decision-making process of customers in selecting, purchasing and utilising the product or service. It is a complicated process and is affected by diversified factors. Previous big data research has attempted to understand consumers' web and mobile usage, social media engagement, and electronic word-of-mouth (eWOM) behaviour, as well as the behavioural effects of online communities and social networks.

Customer web/mobile usage behaviour. Research on this topic involves extensive use of web and mobile analytics. One line of research examines customers' mobile usage behaviour. For example, Ghose and Han (2011) find that the geographical mobility of users and social networks have positive influence on mobile Internet usage behaviour while multimedia content generation has a negative influence. Ranking effects (Ghose, Goldfarb, & Han, 2013) and rewarding (Clausse, Kretschmer, & Mayrhofer, 2013) promote user engagement and mobile app success, but users' preferences and consumption of mobile apps are significantly diverse (Han, Park, & Oh, 2016). Nevertheless, Xu, Forman, Kim, and Van Ittersum (2014) find that the adoption of mobile apps can increase corresponding mobile website visits. Moreover, web and multimedia data are used in a way to investigate online learning and web browsing behaviour. For example, Mayzlin and Yoganarasimhan (2012) discover that blogs with links to other blogs have high news-breaking ability, leading to more readers and enhanced learning. He (2013) uses video stream data to identify patterns in online learning behaviour. Ding, Li, and Chatterjee (2015) document evidence that concurrent learning of users' behaviour is beneficial to real-time, intent-based optimal interventions, which increases purchase likelihood.

Table 2-1. Marketing perspective of big data

Marketing perspective		Selected studies
Consumer behaviour	Customer web/mobile usage behaviour	Claussen, Kretschmer, & Mayrhofer (2013); Ding, Li, & Chatterjee (2015); Ghose & Han (2011); Ghose, Goldfarb, & Han (2013); Han, Park, & Oh (2016); He (2013); Mayzlin & Yoganarasimhan (2012); Singh, Sahoo, & Mukhopadhyay (2014); Xu, Forman, Kim, & Van Ittersum (2014)
	Customer engagement with social media	Amaro, Duarte, & Henriques (2016); Guesalaga (2016); Harrigan, Evers, Miles, & Daly (2017); Oh, Roumani, Nwankpa, & Hu (2017); Pagani & Malacarne (2017); Van Doorn et al. (2010); VanMeter, Grisaffe, & Chonko (2015)
	Electronic word-of-mouth	Balaji, Khong, & Chong (2016); Cantallops & Salvi (2014); Cascio et al. (2015); Eisingerich, Chun, Liu, Jia, & Bell (2015); Hennig-Thurau et al. (2015); King, Racherla, & Bush (2014); Lee & Song (2010); Mayzlin (2006); Sonnier et al. (2011)
	Online community	Chau & Xu (2007); Garg, Smith, & Telang (2011); Johnson, Safadi, & Faraj (2015); Lu et al. (2013); Ludwig, De Ruyter, Mahr, Wetzels, Bruggen, & De Ruyck (2014); Wang et al. (2013)
	Social network effect	Fang, Hu, Li, & Tsai (2013); Goh et al. (2013); Rapp et al. (2013); Shen, Chiou, Hsiao, Wang, & Li (2016); Shriver, Nair, & Hofstetter (2013); Wu (2013); Zhang, Li, Burke, & Leykin (2014)
Consumer sentiment	Online ratings	Ghose, Ipeirotis, & Li (2012); Hu, Bose, Koh, & Liu (2012); Lee et al. (2015); Moe & Trusov (2011); Sridhar & Srinivasan (2012); Sun (2012)
	Online reviews	Baek, Ahn, & Choi (2012); Büschken & Allenby (2016); Cao, Duan, & Gan (2011); Chen & Xie (2005); Cheng & Ho (2015); Goes, Lin, & Au Yeung (2014); Ludwig, De Ruyter, Friedman, Brüggen, Wetzels, & Pfann (2013); Moon, Park, & Kim (2014); Ngo-Ye & Sinha (2014); Singh, Irani, Rana, Dwivedi, Saumya, & Roy (2017); Sparks & Browning (2011); Vermeulen & Seegers (2009)
	Sentiment analysis	Alfaro et al. (2016); Chen, Zheng, & Ceran (2016); Das & Chen (2007); Gao, Greenwood, Agarwal, & Jeffrey (2015); Gopaldas (2014); Hildebrand, Häubl, Herrmann, & Landwehr (2013); Homburg, Ehm, & Artz (2015); Kang & Park (2014); Mostafa (2013); Ordenes et al. (2014); Xiang, Schwartz, Gerdes, & Uysal (2015); Zhan, Loh, & Liu (2009)

Marketing strategy	Brand analysis	Beukeboom, Kerkhof, & De Vries (2015); Camiciottoli, Ranfagni, & Guercini (2014); Culotta & Cutler (2016); Godey, Manthiou, Pederzoli, Rokka, Aiello, Donvito, & Singh (2016); Gretry, Horváth, Belei, & van Riel (2017); Költringer & Dickinger (2015); Moro, Rita, & Vala (2016); Nam & Kannan (2014); Nguyen, Yu, Melewar, & Chen (2015); Tirunillai & Tellis (2014)
	Digital and social media marketing	Ghose & Todri (2016); Gopinath, Chintagunta, & Venkataraman (2013); Iyer & Katona (2016); Järvinen & Karjaluoto (2015); Kumar, Choi, & Greene (2017); Li, Lin, & Chiu (2014); Liu & Mattila (2017); Malthouse, Haenlein, Skiera, Wege, & Zhang (2013); Mariani, Di Felice, & Mura (2016); Trusov, Ma, & Jamal (2016)
	Market analysis	Archak, Ghose, & Ipeirotis (2011); Chong, Li, Ngai, Ch'ng, & Lee (2016); France & Ghose (2016); Lee & Bradlow (2011); Leeflang, Verhoef, Dahlström, & Freundt (2014); Nassirtoussi, Aghabozorgi, Wah, & Ngo (2014); Netzer et al. (2012); Schneider & Gupta (2016); Thorleuchter & Van den Poel (2012)
	Mobile marketing	Andrews et al. (2016); Fong et al. (2015); Ghose & Han (2014); Grewal, Bart, Spann, & Zubcsek (2016); Li & Du (2012); Luo et al. (2014); Shankar, Kleijnen, Ramanathan, Rizley, Holland, & Morrissey (2016)
	Recommendation and personalisation	Chung et al. (2016); Colace, De Santo, Greco, Moscato, & Picariello (2015); Feng, Tian, Wang, & Li (2015); García-Cumbreras, Montejo-Ráez, & Díaz-Galiano (2013); Ho et al. (2011); Hyung et al. (2014); Lee (2007); Oestreicher-Singer & Sundararajan (2012); Rust & Huang (2014); Yang et al. (2008)
	Social media management	Dellarocas (2006); Dijkmans et al. (2015); Godes & Mayzlin (2009); Gu & Ye (2014); Hamilton, Kaltcheva, & Rohm (2016); Ibrahim, Wang, & Bourne (2017); Kim, Lim, & Brymer (2015); Kumar, Bezawada, Rishika, Janakiraman, & Kannan (2016); Miller & Tucker (2013); Risius & Beck (2015); Tripp & Grégoire (2011); Van Noort & Willemsen (2012)

Customer engagement with social media. Another line of research focuses on customer engagement behaviour in social media and its business impact. VanMeter et al. (2015) conceptualise attachment to social media to explain customers' social media activities. Amaro et al. (2016) study travellers' use of social media for travel purposes, and likewise Harrigan et al. (2017) attempt to measure how customer engage with tourism brands on social media. Guesalaga (2016) discovers that the use of social media can benefit B2B selling and that this can be partly predicted based on customer engagement with social media. Oh et al. (2017) also find economic benefits of online consumer engagement behaviour on social media. Given the potential impact of customer engagement, the factors that drive personal engagement and social interactive engagement have also been explored by a number of studies (e.g., Pagani & Malacarne, 2017; Van Doorn et al., 2010).

Electronic word-of-mouth. The current literature shows a considerable interest in eWOM (King et al., 2014). Two main lines of research have emerged, which are concerned with the antecedents and consequences of consumers' eWOM behaviour (Cantallop & Salvi, 2014). In the research stream on eWOM antecedents, a number of studies have explored the determinants of eWOM behaviour. For example, Lee and Song (2010) pay special attention to online complaints and find that informational factors such as vividness and consensus make an impact on customers' evaluation of businesses. Similarly, Balaji et al. (2016) find negative eWOM may be attributed to a variety of factors such as tie strength and feelings of injustice. Eisingerich et al. (2015) argue that eWOM behaviour on social media sites is related to social risk, which may lead to less willingness to participate in eWOM. Regarding the consequences of eWOM, there is substantial evidence that eWOM has significant influences on customers' purchase intentions, sales, and post-purchase eWOM evaluations and recommendation behaviour (e.g., Cascio et al., 2015; Hennig-Thurau et al., 2015; Sonnier et al., 2011).

Online community. The increasing interaction via the Internet is breeding online communities, which are virtual communities where members acquire information and communicate with each other through social network platforms. Current studies are focused on detecting online communities as well as identifying characteristics within those communities. Several studies have proposed methods to detect groups in virtual communities (e.g., Chau & Xu, 2007; Wang et al., 2013), discover information (e.g., Garg et al., 2011), and identify communities (e.g., Ludwig et al., 2014). Besides, there are discussions on emerging online community leadership and identification, especially linguistic style matches, which shape community dynamics (Johnson et al., 2015) and drive network growth (Lu et al., 2013).

Social network effect. Consumer behaviour is also influenced by social networks, where the pattern and dynamics, and influencing entities may have a great impact. Prior research has documented substantial evidence of such an impact. For example, purchase behaviour is affected by online social media brand communities (e.g., Goh et al., 2013) and interactive social influences (e.g., Zhang et al., 2014). Adoption decisions and probabilities can be predicted by analysing social networks (Fang et al., 2013). Shriver et al. (2013) find that online UGC has a positive relation with their social ties and that the network effects can boost advertising and revenue growth. Besides, social media can enrich network information, which has a positive effect on work productivity and job security (e.g., Wu, 2013), brand and retailer performance as well as consumer–retailer loyalty (e.g., Rapp et al., 2013).

2.4.4.2 Consumer sentiment

Consumer sentiments reflect consumers' feelings, perceptions and evaluations of products or services. In this research stream, sentiment analysis is a hot topic and there have been advancements in analytics techniques and wide application. This approach has been applied to detect consumers' opinions and assess customer satisfaction with specific commodities to understand their consumption experience (Dellarocas, Zhang, & Awad, 2007). Besides, social media including online reviews and rating systems have been designed and studied in a way to generate positive outcomes. One point to note here is that studies on big consumer data are not limited to marketing purposes. Looking into consumer opinions can also shed light on operation and production improvement.

Sentiment analysis. Sentiment analysis extracts and classifies subjective information in various data sources, which can be applied to improve business intelligence. A synonym, opinion mining, often refers to the same field of study. Overall, sentiment analysis provides useful information for decision-making (Alfaro et al., 2016). For example, marketplace sentiments can advance companies' understanding of consumer satisfaction and experience (e.g., Xiang et al., 2015; Gao et al., 2015), which is beneficial to niche market identification (Gopaldas, 2014) and brand positioning (Mostafa, 2013). Web comment text, social media messages (e.g., microblog), product/service reviews and other types of UGC are commonly used in these studies. In addition, sentiment detection and classification as part of sentiment analysis have also attracted research interest. Based on practical purposes, a number of new methods have been explored to detect emotions (e.g., Balahur et al., 2012; Das & Chen, 2007; Gao, Xu, & Wang, 2015), spot topics (e.g., Li & Wu, 2010; Zhan et al., 2009), and improve sentiment classification accuracy (e.g., Colace et al., 2015; Da Silva et al., 2014; Khan, Bashir, & Qamar, 2014). They are broadly applied to analyse the sentiments and opinions of consumers and markets, so as to enhance overall management efficiency.

Online ratings. Consumers evaluate products or services by giving numerical scores and prior research has focused on the impact of online ratings. For example, Moe and Trusov (2011) illustrate that online product ratings dynamics have direct and immediate effects on sales. Sun (2012) further proves that a high variance in product ratings can help increase sales only if the average rating is low. Besides, online ratings have a social influence on other users' ratings behaviour (Lee et al., 2015). Sridhar and Srinivasan (2012) find that the ratings offered by peer consumers can moderate the effects of product experience on customers' ratings. To leverage online ratings, Ghose et al. (2012) propose a ranking system that analyses users' ratings to assess customers' preferences, hence providing the best-fit product and service. Moreover, a few studies (e.g., Hu et al., 2012) have found that firms may manipulate product ratings and this requires further attention from business operators.

Online reviews. Online reviews are a form of e-WOM communications that have a significant impact on consumers' choices and transaction or non-transaction related decisions. For example, Ludwig et al. (2013) find that the textual content and linguistic style of reviews drive changes in online conversation rates. Vermeulen and Seegers (2009) prove that the existence of online reviews positively affects consumers' awareness and thus consideration of service providers. Sparks and Browning (2011) find similar evidence of the favourable impact of online reviews on customers' purchase intentions. Accordingly, it is important to improve the helpfulness of reviews. The extant research has explored the influence of review helpfulness and tried to predict helpfulness by looking into text linguistic features or reviewer engagement characteristics (e.g., Baek et al., 2012; Cao et al., 2011; Cheng & Ho, 2015; Ngo-Ye & Sinha, 2014; Singh et al., 2017). In addition, to enable a deeper analysis of customer reviews, advanced text-mining approaches have been developed and improved, drawing on language features, thematic structure or semantic information.

2.4.4.3 Marketing strategy

The above-mentioned big customer data facilitates in-depth analysis of the market and brand. This advances marketing strategies, such as mobile targeting, social media marketing, personalisation and recommendation. More importantly, predictive and prescriptive analytics examine real-time marketing performance and enable dynamic adjustments of strategies (Nichols, 2013).

Brand analysis. In general, brand analysis pins down brand position in the market, brand perception by consumers, competitors' brand performance, and so forth. Big data brand analysis is primarily conducted in a social context. For example, online information has an influence on consumers' perceptions of brands. Camiciottoli et al. (2014) find consistent brand associations in an online community of international consumers. Tirunillai and Tellis (2014) demonstrate that dynamic analysis of online UGC can reflect consumer satisfaction, thus improving competitive brand position. Nam and Kannan (2014) suggest the implications of social tagging for brand performance measurement and brand equity management. Schweidel and Moe (2014) point out that there are differences across multiple social media venue formats, which may have a potential influence on brand sentiment. Nguyen et al. (2015) demonstrate that social media strategic capability can enhance brand innovation.

Market analysis. There are several directions in leveraging big data in market analysis. One is market prediction through mining textual and web information from company websites (e.g., Nassirtoussi et al., 2014). It has been proven that such information is useful in predicting commercial success (Thorleuchter & Van den Poel, 2012). The second topic is about sales predictions based on UGC and customer sentiments. For example, Archak et al. (2011) find that the textual content of product reviews has a significant impact on consumers' choices and that mining review text can reveal customers' preferences, which can be incorporated into consumer choice models to estimate future sales changes. Chong et al. (2016) also find that online reviews and sentiments can help predict future sales performance. In addition, another area highlights the use of UGC in enhancing marketing efficiency. It has been discovered that UGC can be applied to improve the mapping of market structure (e.g., Netzer et al., 2012), detect customer–website interactions (e.g., Schäfer & Kummer, 2013), and identify future profitable customers more accurately (e.g., D'Haen, Van den Poel, & Thorleuchter, 2013;

Thorleuchter, Van den Poel, & Prinzie, 2012). In particular in e-commerce, by capturing detailed customer behaviour information, the knowledge management strategy in marketing can help companies gain competitive advantage in business activities through establishing better interpersonal relations with customers, suppliers, business partners and employees. Digital data is playing an increasingly important role in B2C and B2B marketing, but there are also challenges facing companies that need to be further addressed (Leeflang et al., 2014).

Mobile marketing. A significant interest in this line of research is targeted advertising. In recent research, mobile targeting and advertising have been proved to be effective for location-based services (Li & Du, 2012). Mobile advertising allows for targeting segmented consumers and communicating with them based on specific behavioural, demographic, psychographic and other features so as to provide relevant offers (Grewal et al., 2016). By incorporating locational and geographical parameters, retailers have more power in offering discriminated prices to increase sales (e.g., Fong et al., 2015; Luo et al., 2014). Andrews et al. (2016) illustrate that physical crowdedness has positive impacts on consumers' responses to mobile ads, which is beneficial to hyper-contextual mobile advertisements. Nonetheless, social effects on advertising may vary across markets with different demographic characteristics and groups (Gopinath et al., 2013).

Digital and social media marketing. Research on this topic has predominantly investigated the effectiveness of digital advertising and targeting. For example, Li et al. (2014) examine customer corresponding click-through rates in an advertisement campaign by leveraging social context and social influence. Mariani et al. (2016) study the use of Facebook in promoting tourism destinations and identify that the visual content and length of posts have a positive influence on user engagement. Ghose and Todri (2016) explore the effects of online advertising and find that it is an effective approach to increasing the probability of consumers searching for and purchasing the product or service. Kumar et al. (2016) argue that social media marketing and traditional marketing can create synergistic effects and the effectiveness of these marketing efforts varies over time, which should be taken into consideration in integrated marketing communications.

Social media management. Personal interactions over social media are becoming more frequent. A review of the relevant literature shows that firms tend to use social media as a platform to promote and deliver information to customers and that this firm-generated content has a significant influence on consumers' perceptions and decisions. For example, a few studies (e.g., Gu & Ye, 2014; Van Noort & Willemsen, 2012) have explored firms' participation in online communication and documented some evidence of the effectiveness of this for improving customer satisfaction. Kumar et al. (2016) investigate the online marketing communications created by firms in social media and find a positive impact on customers' buying behaviour and hence profitability. Ibrahim et al. (2017) discover that conversations established between online retailers and customers on social media like Twitter affect customers' sentiments and evaluation of brands. Besides, prior studies have also discussed the potential benefits of active management of social media in enhancing customer engagement. For instance, Miller and Tucker (2013) find incremental engagement with social media of clients and employees when organisations actively manage their social media presence. Dijkmans et al. (2015) demonstrate that corporations' online activities encourage consumers' social media use, which can enhance corporations' reputation, especially among non-customers. This line of research stresses the importance of developing effective social media management strategies to mobilise massive activities and information on online social platforms.

Recommendation and personalisation. In the Web 2.0 era, recommendations are becoming more customised. UGC and the sentiments therein are being analysed to accommodate customer needs (e.g., Colace et al, 2015; García-Cumbreras et al., 2013; Hyung et al., 2014). Personalised recommendations are achievable with technological improvements and big customer data (Rust & Huang, 2014). According to the findings in Brown et al. (2011), advanced analysis and customisation are attainable with the use of real-time and wide-ranging data streams. Through routing location (e.g., Yang et al., 2008), social network (e.g., Chung et al., 2016), community (e.g., Feng et al., 2015), and personalised information (e.g., Fan et al., 2006), user preferences and behaviour can be detected and predicted, which can promote personalisation in marketing to a higher level.

To conclude, from the marketing lens, customers are the priority and understanding their behaviour and opinions is the primary concern for marketing researchers. Nevertheless, marketing practice should be integrated into a higher level strategic framework with an organisational commitment to a data-driven culture. This leaves research space for exploring firms' strategic actions in regard to their digital marketing campaigns and their engagement in online communication activities, so as to play a role in online social interactions and influence the networks.

2.4.5 Other perspectives of big data

Apart from the above main streams in the current big data literature in the management field, a number of researchers have looked into big data from several other perspectives.

First, in the *accounting and finance* area, big data is regarded as an informative source that can affect and predict firm financial performance (Balakrishnan, Qiu, & Srinivasan, 2010). In particular, social media has been shown to have a great effect on stock performance (Schniederjans, Cao, & Schniederjans, 2013; Yu, Duan, & Cao, 2013). Besides, by incorporating big data, the nature of accounting and audit judgements are also changing (Brown-Liburd, Issa, & Lombardi, 2015; Vasarhelyi, Kogan, & Tuttle, 2015). Textual data is being effectively used to detect financial fraud in reports (Glancy & Yadav, 2011). Cao, Chychyla, and Stewart (2015) discuss the potential adoption of big data analytics to improve financial statement audit efficiency. Although the opportunities and potential benefits of big data are significant for the accounting and finance domain, the challenge is to combine traditional and non-traditional data sources to support accounting and finance practices.

Second, there is the *international business* perspective, which refers to the big data driven studies on the performance of trade and investment activities by firms across national borders (see Cavusgil, Knight, & Riesenberger, 2012). In the Web 2.0 environment, big data is expected to contribute to global business. As illustrated in Lau, Liao, Wong, and Chiu (2012), online environment scanning using Web 2.0 helps improve decision-making regarding cross-border mergers and acquisitions. Okazaki and Taylor (2013) focus on the use of social media for international advertising and identify several theoretical foundations for future search. There remains research space around this topic to explore the potential use of big data in international business decisions and operations.

Furthermore, in *public administration,* big data has been applied in several areas. For example, one recent study has focused on how big data could be utilised in combatting health emergencies such as Ebola (Amankwah-Amoah, 2016). Likewise, Panagiotopoulos, Barnett, Bigdeli, and Sams (2016) discuss the use of social media in communicating risks to the public and find that it is a helpful tool in emergency management. Moreover, e-voting and e-government can benefit from big data (Zissis & Lekkas, 2011). Suh, Park, and Jeon (2010) discover that text mining can be combined with data mining to efficiently detect and forecast petition trends in e-government. In contrast, Huberty (2015) argues that social media can be used to reach prospective voters but it is not a reliable source to assess public sentiments and forecast elections. Nevertheless, Grimmelikhuijsen, and Meijer (2015) find that due to greater transparency and participation, the use of social media can enhance perceived police legitimacy and big data may assist in improving the management of public programmes (Lavertu, 2015).

2.4.6 Summary of big data research in management

The foregoing literature review highlights the key interest in the current big data research in the field of business management. First, the papers in the information management field deal mainly with analytics techniques and technological issues. The advanced approaches help improve information retrieval and data processing due to better accuracy and efficiency, and therefore support research needs in other management fields. Second, the organisational view of big data emphasises its impacts on strategic decision-making and the changes it brings to organisational structure, functions and management. The use of big data is leading to novel and superior strategies in marketing, operations and other areas with attempts to enhance business efficiency and performance in the digital business environment. Third, researchers in the operations domain have discussed big data in terms of production, innovation, and supply chain management so as to enhance operational efficiency and effectiveness by leveraging the real-time information delivered by big data. In addition, big data is attracting extensive interest from marketing scholars. Understanding consumers and thus developing effective marketing strategies seem to be the primary directions in the extant research. These studies analyse textual data, social media data, and mobile and sensor data to uncover the consumer sentiments and behaviour hidden in user-generated information, taking into account social influence and the specific nature of the online interactive environment.

These findings show that considerable attention is being paid to big data's potential for creating value so as to advance research and management in the data-rich future. The mounting body of research across disciplines indicates that the capability of the data-driven approach is not only evident at the technical stage in extracting and processing information, but can also promote changes in managing organisations, operations, marketing and other business activities. By leveraging big data in the management mechanism, additional value can be discovered, created and realised in regard to business development. However, even though management professionals and researchers have a growing awareness of big data, the full benefits have not yet been accrued as a result of the rapid changing environment in relation to technological advancement. In the process of big data value achievement, all management activities are connected and contributing. To unlock the full value, firms need to formulate and implement a data-driven strategy. Top management teams should also make strategic adjustments within their organisations through measures such as investment in IT innovation and data analytics skills development. Accordingly, decisions would be made with enhanced flexibility and accuracy by applying advanced analytics to analyse and understand massive data. The overall performance of operations, marketing, and other business activities principally depend on the quality of strategic decision-making, which also determines the realisation of profits and competitive advantage.

The literature review reveals great needs and space for big data research in the realm of business management. First, theoretical development is the basis of providing general guidance for researchers to implement big data methods. It is also worth noting that 'technology transfer' is critical for researchers to identify the best use of analytics techniques to achieve optimal performance. Moreover, a number of fruitful areas, as outlined in Table 2-2, may potentially benefit future big data research development in management communities. The prospect of big data value underscores the importance of strategising about big data throughout the management ecosystem to seize the initiative.

Table 2-2. Perspectives and future directions of big data research

Perspective	Key interest	Big data impact	Research context	Avenue for future research
Information management	Advanced analytics technique application	· Challenges in data management and analysis · Advanced algorithms	· Data acquisition and process · Modelling	· Models and techniques evaluation and comparison · Data quality improvement
Organisation	Organisational change and strategic decision-making	· Data-driven decision · Ecosystem change · Management process	· Organisation structure · Organisation functions · Big data commitment · Business intelligence · Strategic action	· Organisation alignment with data-driven strategy · Human resource management improvement · Reshape of strategic operational mechanism and actions · New perceptions of the value of experience, nature of expertise, the practice of management
Operation	Operation efficiency	· Real-time control · Reduced cost and risk · Innovation	· Production, innovation and supply chain · Sensor/web/social media	· Innovation e.g., new product and operation process · Further improvement of operation efficiency · Big data supply chain evaluation
Marketing	Consumer and marketing effectiveness	· Online evaluation · Personalisation · A better understanding of consumers and market	· Social media/web/sensor · Advertising · Social influence · Consumer behaviour and sentiment	· Firm engagement with online communication · New strategy evaluation · Possible adverse effect of online social network marketing · Segmented pricing and revenue management
Others	Accounting and finance, Public sector, International business	· Firm performance · Greater transparency and participation · A better understanding of foreign market	· Social media · Management actions · Cross-border decision and operation	· Accounting quality · Challenges and risks of big data use in public sector · International decision and business operation (e.g., marketing, supply chain, merger and acquisition, investment)

2.5 Strategising about Big Data

This section reviews the relevant literature on big data value in deriving competitive advantage. The two opposing views on big data are first presented, followed by the challenges discussed by prior research. Then the opportunities of big data in creating business impacts are demonstrated by surveying the current body of management studies to identify ways of utilising diversified types of big data. Finally, drawing on the dynamic capability view, the prominence of strategising about big data from the online crowd in creating business value is further discussed.

2.5.1 Competing schools of thought on big data

The accumulated body of literature suggests two perspectives on big data effects: big data as an asset and big data as a liability. On the one hand, the big data as an asset perspective (Bughin, Chui, & Manyika, 2010; Bughin, Livingston, & Marwaha, 2011; Economist Intelligence Unit, 2011; Perrons & Jensen, 2015) contends that rooted in big data are ample opportunities, which can be utilised by businesses to outwit their rivals through developing proactive strategies for future success. Recent studies indicate that, by mobilising big data, firms are able to gauge customers' opinions and take steps to improve their experiences, deliver personalised products and services, enhance internal efficiency and operations, and develop a better pricing strategy (Li & Wang, 2017; Alharthi, Krotov & Bowman, 2017; Lee, 2017). Specifically, from the strategy and marketing perspectives, one big data impact revolves around digital brand touch points (i.e., contact between a potential customer and a firm brand) via social media, the Internet and mobile devices (Malthouse & Li, 2017). Unlike traditional offline touch points such as viewing a printed advert in a physical magazine, digital brand touch points can easily be monitored and recorded on a large scale to produce big data on consumers (Malthouse & Li, 2017). Accordingly, organisations are then able to harness such data to inform advertising decisions, targeted marketing campaigns and branding strategies. Another widely elaborated effect of big data is its ability to help predict and take preventative measures to deal with or address organisational problems such as consumers concerns and complaints in a timely manner. Improvements in modern technologies have made it increasingly easier to track individuals' behaviour, experiences and habits (Lee, 2017). A wide array of publicly available social interaction information is brought into the decision process of firms in tandem with traditional sources of data. The positive view incorporates the various synergistic benefits of big data and suggests that the ability to utilise big data is one of the key sources of competitive advantage for firms (Alharthi et al., 2017; Amankwah-Amoah, 2015).

On the other hand, recent streams of research have demonstrated that data sharing is a cornerstone of how firms strategise across industries and sectors. Although the collection of vast amounts of electronic data from customers, clients and citizens may represent an opportunity to create business and social benefits, it also has the potential to infringe individuals' privacy (Craig & Ludloff, 2011; Goldfarb & Tucker, 2011; Lee, 2017). In this regard, the second perspective on big data, which sees it as a liability (O'Neil, 2016), contends that privacy and data security are increasingly under threat due to the improved capabilities to access individual data. Studies demonstrate that there are increasing opportunities for firms and governments to assemble multiple data on citizens and invade their privacy and that Internet businesses can store and share user-generated data with government agencies, which can also lead to infringements of individual privacy (Oram, 2014). Researchers have also observed that privacy concerns or risks may stem from data security breaches in different organisations (Craig & Ludloff, 2011). Although there has been a surge in providing individuals with options to opt in or out of data collection, many organisations have a tendency to keep individuals' personal data and the process of requesting deletion can be protracted (Perri, 2002). To a greater extent, this view sees little value in the widespread collection of individuals' data for profit or surveillance purposes. This line of research documents various negative effects of using big data to predict and identify patterns of human behaviour. It has also criticised big data applications for lacking the emotional element inherent in how humans make decisions and communicate. The inability to capture these features could lead to the misinterpretation of data and consequently incorrect decisions. Such decisions could lead to the misallocation of resources and the misdirection of managerial attention. Tangentially related to this line of thinking is the suggestion that the promising opportunities of big data are 'hype' rather than reality (Franks, 2012).

Deviating from these two competing views, another view is a "middle-of-the-road" perspective, which focuses on minimising the extremes associated with the two competing perspectives on big data. There are some underlying guidelines rooted in this moderate view. First, in seeking to capture and utilise big data, firms should concurrently respect the privacy of individuals to help minimise the negative effects (Christiansen, 2011). In addition, there is a need for organisations to respect the local sensitivity of the data collected. This guiding principle should curtail or shape how the data is utilised. When stakeholders from the public or private sectors exercise their big data analytics capabilities to improve organisational performance, they should also develop an appropriate big data strategy to preserve individuals' privacy and address security concerns. This includes using data protection to help protect individuals from unnecessary intrusion and ensuring individuals' ability to preserve their personal data.

2.5.2 Challenges

Accruing the full benefits of big data is predicated on organisations' ability to overcome certain challenges. First, as previously indicated, preserving individuals' digital footprints from misuse remains a major issue in harnessing big data in the digital era. Some data is of a personal or sensitive nature (Alharthi et al., 2017; Douglas, 2013) and it is incumbent upon data users to ensure that privacy is not infringed in the quest to capture value from big data. This suggests the need for firms to capture the positive externalities of big data whilst minimising the negative effects, including infringement of civil rights and individuals' privacy. This requires organisations to go beyond their own interests and to consider more broadly societal improvements through big data research.

Besides, as discussed in Section 2.3, the complexity inherent in the nature of big data challenges efficient data management and meaningful analysis. Recent works have underscored that one of the challenges is to meet the growing demand of big data specialists, including not only data scientists, IT specialists, and mathematicians, but also strategists, economists and sociologists (Lee, 2017). Investment in upgrading managers' and analysts' skills and expertise in advanced analytics is required to tackle the growing need for businesses to make data-driven decisions (Manyika et al., 2011).

In addition, although scholars and practitioners have long touted the benefits that can be accrued from big data, firms are often confronted with difficulties in justifying investments in big data (Lee, 2017). Some organisations have abstained from investing scarce resources to harness big data largely due to the perception of the top management team that there is "little value in pursuing big data initiatives" (Alharthi et al., 2017, p. 289). Accordingly, such organisations may require a cultural change to help ensure efficient alignment between the internal structures and changes in the external environment, including the introduction of new technology and analytics to work with big data.

2.5.3 Opportunities for utilising big data

As previously noted, big data offers exciting opportunities to businesses. It has the potential to play a significant role in countries' development, academic research, and the way people see the present and future worlds (Jin et al., 2015). There is strong evidence indicating that big data is an essential element for firms seeking competitive advantage (Brown et al., 2011). It can be harnessed to inform and improve decisions, and less decisions are now being made in a vacuum or based on intuition, but, rather, are based on solid data. Data in numerous formats including structured, unstructured, and semi-structured data is being generated every second, informing decision makers about businesses with new insights and knowledge. Mobilising big data provides a source of sustainable competitive advantage based on close collaboration among various stakeholders to share knowledge and experience in the value creation process. Indeed, the potential of big data for making a positive influence on business, such as in enhancing marketing strategies, strengthening customer relationships, lowering management risks, designing new products and services, improving operation efficiency and responding to consumers' needs, has been widely acknowledged (Kiron & Bean, 2013). The current literature has started to utilise big data, in particular unstructured data, and advanced analytics in a number of ways to improve business performance. Based on the previous review of big data research in business and management (Section 2.4), Table 2-3 summaries the emerging applications and purposes of leveraging various types of big data in business and management.

Table 2-3. Selected research on big data utilisation

Data	Emerging applications	Purposes	Selected studies
Text	Customer behaviour analysis	Use online customer reviews to evaluate customer satisfaction, predict purchase behaviour and sales, understand customer engagement/eWOM behaviour, assess perception and peer influence in the community.	Archak et al. (2011); Cheng & Ho (2015); Goes et al. (2014); Gu & Ye (2014); Kang & Park (2014); Ludwig et al. (2013); Moon et al. (2014); Ordenes et al. (2014); Sparks & Browning (2011); Vermeulen & Seegers (2009); Zhan et al. (2009); Zhu & Zhang (2010)
	Market analysis	Use narrative disclosure to evaluate risks, predict market performance, detect brand association and market structure, and influence financial management.	Balakrishnan et al. (2010); Bao & Datta (2014); Camiciottoli et al. (2014); Das & Chen (2007); Glancy & Yadav (2011); Költringer & Dickinger (2015); Nassirtoussi et al. (2014); Nassirtoussi et al. (2015); Netzer et al. (2012); Thorleuchter & Van de Poel (2012)
	Opinion mining	Use online comments and other UGC to detect sentiment polarities and topics, predict opinion trends, discover knowledge and make recommendations.	Alfaro et al. (2016); Fan, Wallace, Rich, & Zhang (2006); García-Cumbreras et al. (2013); Ghose et al. (2012); Hyung et al. (2014); Özyurt & Köse (2010); Tang & Guo (2015); Ur-Rahman & Harding (2012); Ye, Zhang, & Law (2009); Yoon (2012)
Mobile, sensor, and web	Online and mobile marketing	Use users' locations, preference, social network and other personalised information collected from customers' interactions with web and mobile devices to detect user interests, provide personalised service and recommendation, and predict profitability.	Andrews et al. (2016); Chung et al. (2016); Feng et al. (2015); Fong et al. (2015); Ho et al. (2011); Järvinen & Karjaluoto (2015); Li & Du (2012); Luo et al. (2014); Schäfer & Kummer (2013); Thorleuchter et al. (2012); Wang et al. (2013); Yang et al. (2008); Yeh et al. (2009)
	Operation management	Use real-time data generated from web, mobile and sensors to support operation, such as order tracking and prediction, learn user behaviour in real-time, optimise supply chain and dynamic pricing.	Castellanos et al. (2012); Ding et al. (2015); Guo et al. (2014); Huang & Van Mieghem (2014); Li & Wang (2017)

Network	Online marketing	Use social media in advertising, recommendation, customer relationship management, brand evaluation and innovation, sentiment analysis, sales prediction, and market analysis.	Colace et al. (2015); Godes & Mayzlin (2009); Gopaldas (2014); Gopinath et al. (2013); Lee & Song (2010); Li et al. (2014); Malthouse et al. (2013); Mayzlin (2006); Moe & Trusov (2011); Nam & Kannan (2014); Nguyen et al. (2015); Rapp et al. (2013); Schweidel & Moe (2014); Sun (2012);
	Online community and social influence	Use social media and online network data to detect online community, understand peer and network effects, predict and affect consumer behaviour.	Beukeboom et al. (2015); Cascio et al. (2015); Chau & Xu (2007); Fang et al. (2013); Goh et al. (2013); Johnson et al. (2015); Lu et al. (2013); Homburg et al. (2015); Iyer & Katona (2016); Oestreicher-Singer & Sundararajan (2012); Shriver et al. (2013); Sridhar & Srinivasan (2012); Van Noort & Willemsen (2012); Zhang et al. (2014)
	Organisation management	Use social media in human resource practice, firm value prediction, online evaluation, reputation and firm performance enhancement.	Dijkmans et al. (2015); Kluemper & Rosen (2009); Luo et al. (2013); Roth et al. (2013); Schniederjans et al. (2013); Van Iddekinge et al. (2016); Yu et al. (2013)
	Production, operation & innovation	Use social media to discover useful information (such as product defect) & innovative ideas to enhance product quality, operation efficiency, & new product development.	Abrahams et al. (2015); Chan et al. (2016); Chan et al. (2017); Mount & Martinez (2014)

2.5.3.1 Use of textual data

The current literature reveals that the rich textual information contained in UGC such as online comments, weblogs and message boards can be analysed for marketing and operations efficiency enhancement. Firms can be empowered to sense the public perception of the firm by mining opinions from the UGC. In this case, computer-aided approaches have been applied to classify sentiment polarities (e.g., Ye et al., 2009), detect topics (e.g., Özyurt & Köse, 2010; Yoon, 2012), predict opinion trends (e.g., Alfaro et al., 2016; Tang & Guo, 2015) and discover knowledge (e.g., Fan et al., 2006; Ghose et al., 2012). Using the insights from online data has proved to be useful in predicting customers' behaviour. In the current literature, online reviews and comments are widely used to evaluate customer satisfaction (e.g., Gu & Ye, 2014; Kang & Park, 2014) and predict consumer purchasing behaviour and sales performance (e.g., Archak et al., 2011; Moon et al., 2014; Sparks & Browning, 2011; Zhu & Zhang, 2010). In addition, narrative disclosure can be useful in designing strategies for businesses to achieve success. Past studies have documented that online user-generated data can be used to improve product innovation (e.g., Chan et al., 2016), make recommendations (e.g., García-Cumbreras et al., 2013; Hyung et al., 2014), evaluate risks (e.g., Bao & Datta, 2014), predict market performance (e.g., Nassirtoussi et al. 2014, 2015; Thorleuchter & Van de Poel, 2012), detect brand association and market structure (e.g., Camiciottoli et al., 2014; Netzer et al., 2012), and influence financial management (e.g., Das & Chen, 2007; Glancy & Yadav, 2011). These studies have focused on how firms have used online user-generated data to understand consumers, analyse the market and develop strategies accordingly.

2.5.3.2 Use of mobile, sensor and web data

Many scholars have suggested that the advancement of mobile technology and the proliferation of mobile use in people's work and lives have stimulated the generation of massive data (see Abrahams et al., 2015). Access to information has fewer obstacles especially in the aspects of time and location. Given that some of the advertising revenue for traditional forms of advertising such as print media and network television have shifted to new media and online, there are now more opportunities for firms to have better 'visibility' of how their message is performing and the size of the target audience reached (Christiansen, 2011; Czinkota, Ronkainen, Sutton-Brady & Beal, 2011). Accordingly, organisations are now more able to minimise the misallocation of resources in determining their marketing channels. Specifically, mobile data has been used for mobile targeting and advertising; it is used to analyse users' geographical information, usage behaviour, social networks and other personal information to facilitate effective promotions and services (e.g., Fong et al., 2015; Li & Du, 2012; Yang et al., 2008). Such real-time data can also come from sensors, which can support business operations, particularly in regard to improving efficiency in supply chain management and inventory management (e.g., Guo et al., 2014; Li & Wang, 2017). In addition, business analytics has seen the active use of web data. The analysis of web pages, web structure and web logs has been applied to detect user interest (e.g., Feng et al., 2015), design recommender systems (e.g., Wang et al., 2013) and predict profitability (e.g., Thorleuchter et al., 2012). Apart from frontline marketing, web data, such as clickstreams, can be helpful for predicting orders and reducing inventory costs (e.g., Huang & Van Mieghem, 2014), learning user behaviour in real-time (e.g., Ding et al., 2015), and improving operational business intelligence capabilities (e.g., Castellanos et al., 2012).

2.5.3.3 Use of network data

Social media has dramatically changed the way people see the world and communicate with each other. The sheer amount of information generated from social media has been applied in business practice to assist production, marketing, and management activities. UGC in social media can be useful for knowledge discovery (e.g., Abrahams et al., 2015; Chan et al., 2017) and hence enhance operational decision-making. This information, which records online users' behaviour and interactions, can also be used to detect online communities (e.g., Johnson et al., 2015) and evaluate social influence in virtual communities (e.g., Oestreicher-Singer & Sundararajan, 2012; Shriver et al., 2013), so as to predict and affect consumer perceptions and behaviour (e.g., Fang et al., 2013; Goh et al., 2013). In so doing, the effectiveness of online marketing strategies can be enhanced by leveraging social context and social influence (Li et al., 2014). This is potentially achieved through more personalised advertising and recommendations (e.g., Colace et al., 2015), better customer relationship management (e.g., Malthouse et al., 2013), and more precise market targeting. Use of social media not only benefits firms' sales prediction and market positioning, but also has great importance for organisational management. Prior studies have demonstrated that social media has an impact on firm performance (e.g., Schniederjans et al., 2013; Yu et al., 2013) and corporate reputation enhancement (e.g., Dijkmans et al., 2015). It also brings changes in personnel decision-making, such as in assessing potential hires by digging for information from their social media profiles, and evaluating employee performance with workforce data (e.g., Roth et al., 2013; Van Iddekinge et al., 2016).

2.5.4 Dynamic capability and strategic firm engagement

The above discussions suggest the potential of big data for creating strategic value in business and management. Considering the rapid change in the digital and data rich landscape, businesses should plan strategically for big data to develop and sustain competitive advantage. According to the dynamic capabilities view (Teece, Pisano, & Shuen, 1997), firms' competences stem from integrating both internal and external resources and these competences need to be continuously renewed to adapt to the changing environment so as to sustain competitive advantage in the long-run. The dynamic capability framework underlines firms' strategic considerations in seizing opportunities and their ability to reconfigure business competences in the transformation of the market, customers and/or technology (Teece, 2007). In particular, Barrales-Molina, Martínez-López, and Gázquez-Abad (2014) explicate the dynamic marketing capabilities model, which involves absorptive capacity and knowledge management in the process of acquiring market knowledge to renew the organisation. The market knowledge obtained from customers, markets, competitors, the environment and many other actors in the market is a resource that firms can use to develop dynamic marketing capabilities in regard to matching market conditions, innovating new products and services, identifying shifts in consumer behaviour and needs, and developing and renewing strategies (Morgan, 2012; Barrales-Molina et al., 2014).

In the digital era, data is the lifeblood of businesses and a valuable firm-specific asset awaiting strategic exploitation (Bollier & Firestone, 2010). As the comparative advantage theory proposes (Hunt & Morgan, 1995), the creation and efficient use of tangible and intangible resources can yield competitive advantage in terms of marketplace position and thereby superior performance in terms of quality, efficiency, sales, and innovation. It is suggested that data strategising, which harnesses extensive data generated from human interactions and machines, can derive comparative resources and thus competitive advantage. For example, big data may enable businesses to identify and respond to customers' needs, power innovation and new product development, and enhance real-time operational awareness and decision-making (e.g., Wei & Wang, 2010; Antons & Breidbach, 2018). The ability to tease out insights from massive data and strategically manage resources differentiates successful firms from underperforming ones and equips firms for the new competitive game.

The rise of big data promotes businesses to act more proactively in the data-rich, open, participative, and interactive environment. As identified from the previous review of big data research in management, there is an increasing volume of OSIs in the computer-mediated environment. Along with this is the question of how businesses can capitalise on these resources to develop strategic plans. Yadav and Pavlou (2014) classify online interactions into four categories: consumer–firm interactions, firm–consumer interactions, consumer–consumer interactions, and firm–firm interactions. Among the four types of interactions, they find that research on the customer–customer interactions mainly focuses on social networks and UGC and it is a great challenge for business to develop strategies so as to monetise the social interactions among customers. They also discover that there is an emerging stream of research that is concerned with firms' strategies and tactics in terms of interacting and engaging with customers. They view the firm–customer interaction as a promising research area taking into consideration the enhanced information transparency on the Internet and the necessity of firms to capture and manage information about consumers' online activities and interactions.

Nevertheless, firm engagement that exists simultaneously with customer engagement in the OSI network (Wei, Miao, & Huang, 2013) has only recently begun to be discussed and the research on strategic firm engagement remains fragmented. Firm engagement in the context of this doctoral research refers to managerial activities as firms interact with customers and engage in customer–customer interactions. Godes et al. (2005) identify four general action plans in regard to firms engaging in social interactions, which are firms acting as an observer, a moderator, a mediator, or a participant. In these roles, firms can learn from customers' online conversations, foster interactions in the virtual community (e.g., Ryu & Feick, 2007), disseminate information (e.g., Porter & Donthu, 2008; Gu & Ye, 2014), and control online communications (e.g., Voorhees, Brady, & Horowitz, 2006). Ramani and Kumar (2008) claim that superior performance can be achieved if firms are able to orientate their business to interact with customers, gain information from such interactions and leverage it to sustain customer relationship. This suggests a need for firms to manage OSIs through engaging in the network and thereby to disentangle how big data from the crowd of consumers can be leveraged to improve firm competency.

2.6 Summary

This chapter reviews the core streams of literature on big data and big data analytics in the current state of management research. Big data is complex and multidimensional in nature and brings exciting opportunities to business as well as great challenges in regard to data management and analytics to unlock its potential value. The scale and scope expansion of data has accelerated technology advancement with upgraded and novel techniques to handle big data. With the promising developments and technological support, big data is becoming increasingly needed and emphasised in modern business operation. The rise of big data has attracted increasing attention from management scholars who are discussing and exploring ways of accruing big data benefits from the ever-changing data rich environment. Nevertheless, there remains considerable research space to address big data challenges. The literature review in this chapter reveals several gaps in the current body of knowledge.

First, there is a lack of research linking big data with applications in the knowledge economy and management domain. Research on the implications of big data for social science is lagging behind the rapid development of technology and industrial practices. Despite this fact, in the management community, the need for research on big data is becoming more manifest due to the transition of business models and decision-making mechanisms into a data-driven pattern as well as its potential in enhancing the efficiency of management practices. By committing to a data-driven approach, companies are able to respond to customers' needs, improve business performance and compete in the digital market with an extensive amount of data being collected and interpreted. However, our knowledge of how to apply the right techniques to translate data into valuable business insights remains limited. This calls for research to explore and examine big data's strategic value in improving firm performance by performing big data analytics in order to advance our understanding of management theories and practices with new insights from big data.

Second, the proliferation of OSIs presents extensive research opportunities to understand markets and consumers in greater detail, and thus improve business performance. However, less research considers OSIs from a business perspective to explore how firms may potentially use, manage and impact on OSIs. Many efforts of the extant research have been devoted to exploring the significance of OSIs from the consumer perspective, such as the social influence of peer customers in virtual communities and CEB in OSIs. This fails to depict a clear picture of how firms strategically engage in OSIs and what the potential business impacts are of making moves towards proactive online interactions with customers. Accordingly, a line for further research would be to examine the effects of strategic online-based firmcustomer interactions. In this doctoral research, the gap of the strategic intent and consequences of firm engagement in OSIs will be bridged with three empirical studies. Using a large data set of online customer reviews and managerial responses, the three studies in this research will conduct big data analytics to examine the efficacy of firms acting on consumers' comments so as to adapt to the social, participative and interactive business environment, and reinvent their business towards a data-driven path.

In addition, another major research gap identified in the literature review is the insufficient understanding of firm-generated content in online marketing communications. The review suggests that text analytics and social media analytics are gaining popularity in the recent development of big data research. These advanced analytics have been extensively performed on UGC, such as online customer reviews, to examine customers' sentiments and engagement behaviour. In contrast, textual content created by firms has been less studied. We have limited knowledge of businesses' communication styles (e.g., language usage, sentiment, written content) and engagement patterns (e.g., frequency, speed) and their interactive effects on online communication and business performance. To fill this gap, this research (to be specific, Chapter 4 and Chapter 5) attempts to investigate firm-generated content through mining the textual, semantic and behavioural information hidden in businesses' online messages in order to enact and enrich strategies of firm engagement in OSIs.

CHAPTER 3

A DATA ANALYTIC APPROACH TO TRANSFORM USER-GENERATED CONTENT INTO SERVICE IMPROVEMENT

3.1 Introduction

In the landscape of technological changes and temporary advantage (D'Aveni, Dagnino, & Smith, 2010; McGrath, 2013), the traditional approach of firms relying solely on the human and social capital embedded within the firm in tandem with external professionals to develop and improve products or services seems inadequate (Moreau & Herd, 2010; Schreier, Fuchs, & Dahl, 2012; Ulrich, 2007). The proliferation of online review platforms has opened new doors for firms to turn to users in seeking to design and improve products or services (Von Hippel, 2005). Consider a customer who has had a wonderful or an appalling experience of a night stay at a hotel, and the customer might share the experience by telling family/friends face-to-face. In this case, the hotel will not know this customer's service experience unless an after-service survey is conducted, which is costly and inefficient. Alternatively, if the customer goes on Internet sites to share the experience, the praises or complaints about hotel services posted are then viewed and shared immediately by thousands of potential customers across the world, as well as service providers. The vast amount of reviews generated by Internet users not only present detailed service experience and collective opinions from customers but also offer service providers opportunities to make a timely intervention and tailor their services to meet customers' needs.

Indeed, a product or service labelled as designed or improved based on users or customers' comments appears to have wider appeal and possess a better chance of success relative to those developed by the designers inside the focal firms (Poetz & Schreier, 2012; Schreier et al., 2012). Why can the source of product or service design influence consumers' preferences (Fuchs, Prandelli, Schreier, & Dahl, 2013) and bring competitive advantage to business? One possible explanation of this 'user-driven philosophy effect' (Dahl, Fuchs, & Schreier, 2015) is that user-generated ideas are more likely to increase customers' identification with the firm. By capitalising on "shared inventiveness" (Lusch, Vargo, & O'Brien, 2007, p. 11), firms are able to gain quicker and comprehensive knowledge about customers' needs (Hogan, Lemon, & Rust, 2002; Rust & Thompson, 2006) and take more effective actions based on the reflections on customers' responses. By drawing on users' insights to improve product or service quality, firms are able to enhance their ability to attract and retain customers (Fuchs et al., 2013). By empowering existing customers to comment or encouraging them to volunteer ideas for improvements, firms are also able to signal to prospective customers that they are a more consumer-oriented organisation (Crawford, 2008; Fuchs, Prandelli, & Schreier, 2010; Fuchs & Schreier, 2011). By demonstrating the willingness and ability to elicit customers' comments and suggestions, firms may receive a more favourable appraisal from customers (Bharadwaj, Nevin, & Wallman, 2012; Dahl et al., 2015). In spite of these potential benefits of tapping into user-generated data to enhance firms' competitiveness (Bharadwaj & Noble, 2015), our understanding of how user-generated content (UGC) can be harnessed to develop and improve service design and delivery remains limited.

Against this backdrop, the main purpose in this chapter is to examine how firms can utilise UGC to inform decisions on service design and improvement. Specifically, the aim is to discover key issues and concerns in consumers' online reviews and how these themes change over time and vary with customer groups and market segments. Identification of key factors affecting customer satisfaction is critical to service research (Liu, Teichert, Rossi, Li, & Hu, 2017; Sharma, Niedrich, & Dobbins, 1999). In the current study, it is contended that information derived from UGC is valuable to firms in pinpointing areas for service design and delivery. To demonstrate this theoretical contention, a topic modelling approach is applied to over 600,000 online reviews to explore insights into customers' primary concerns in relation to hotel service. The dimensions of consumer-perceived service quality are synthesised and prioritised using a SERVQUAL construct to illustrate their associated importance. Results from the Latent Dirichlet Allocation (LDA) analysis demonstrate that both tangible and intangible aspects of service are important to consumers' perception of service quality, but the significance of these dimensions varies with time, market segments and customer groups.

The study contributes to service and marketing research in several ways. First, an emerging body of research has hinted that big data can be harnessed to improve firms' competitiveness (Bharadwaj & Noble, 2015; Sheng, Amankwah-Amoah, & Wang, 2017), yet few studies attempted to examine the precise mechanisms for identifying key and relevant service attributes.

Prior researches on service quality attributes using customer reviews as data sources mainly focus on patterns of ratings for selected hotel attributes or hotel attributes specified by the travel agencies (e.g., Banerjee & Chua, 2016; Liu et al., 2017; Rhee & Yang, 2015). The current study extends this line of service research by applying a text mining approach to identify the key attributes directly from the user-generated reviews. The exploratory study transforms the vast amount of UGC into a knowledge of customers' perception of service experience. It demonstrates how the textual content of online customer reviews can actually be leveraged in guiding the designing and improving service architecture and deepens our understanding of the complex process inherent in identifying customers' concerns to improve service design and delivery. Thus, this study contributes to the big data and service research (e.g., Mortenson, Doherty, & Robinson, 2015; Tirunillai & Tellis, 2014; Xiang, Du, Ma, & Fan, 2017) by providing insights into how data analytics can be marshalled to capture value from user-generated data in service improvement.

Second, there has seen a growing number of studies using data mining and text mining approaches to support decision-making processes (e.g., Chen, Fan, & Sun, 2015; Liu, Schuckert, & Law, 2018; Ordenes et al., 2014). This work applies the topic modelling technique to synthesise and prioritise service quality dimensions based on real data from consumers. The results of the text analysis reveal key content of customers' online reviews by systematically identifying the most important attributes of customer-perceived service quality. Compared to other research methods, such as focus group and survey, the data is freely available and can cover a certain period. It is less biased as the reviews are mainly based on customers' real experience, which offers concrete evidence on market needs with greater accuracy and efficiency. Besides, based on statistical estimation, the machine learning approach discovers the topics and weights from a large corpus of review text. It is more efficient than manual coding (e.g., Zhou, Ye, Pearce, & Wu, 2014) and enables a more holistic view of the attributes hidden in customer reviews.

In addition, compared to other topic modelling studies (e.g., Guo, Barnes, & Jia, 2017), the topic modelling approach used in this study is more robust in determining the optimal number of topics. It is critical to have an objective and robust process as the chosen number of topics can significantly affect the modelling output. Furthermore, a longitudinal analysis, sentiment analysis, and segment analysis are incorporated in the topic modelling approach, which enable us to gain a better understanding of consumers' changing behaviours over time, the primary concerns of satisfied and unsatisfied customers, and customer expectations in different market segments. From the managerial standpoint, the analysis will help hotel managers better understand customers' needs and concerns. For existing hotels, they can use the approach to evaluate their services and identify areas for service improvement by knowing the main concerning and satisfying aspects of hotel services. For new market entrants, they can design the services to better match the needs of their targeted customers.

3.2 Conceptual Framework

Service system refers to the "value-co-creation configurations of people, technology, value propositions connecting internal and external service systems, and shared information" (Maglio & Spohrer, 2008, p. 18). Service sector particularly views the importance of customers' involvement in improving service delivery mechanisms. Following the service-dominant (S-D) logic (Vargo & Lusch, 2004, 2008), the fundamental source of competitive advantage in the service ecosystem is derived from resources that can be leveraged (Vargo & Lusch, 2004), such as knowledge and skills (Vargo & Lusch, 2008). Value is created through interactive relationships and integrated resources (Vargo & Lusch, 2017; Vargo, Maglio, & Akaka, 2008). Customers can act as ideator, designer and intermediary (Lusch & Nambisan, 2015) and these roles respectively reflect beneficiaries' capability to acquire, integrate and share knowledge and resources to envision, configure and facilitate new service development and service improvement. The involvement of customers in service design and redesign opens new opportunities. For example, firms would be able to identify untapped markets through learning about their customers' requirements and then design new services to fulfil that unmet need (Schreier & Prügl, 2008). By collaborating with customers, firms reduce the chances of service failure and enhance new service development success (Gruner & Homburg, 2000). This implied a compelling need for firms to explore how to utilise customers' insights.

The significance of customers' involvement in service design and improvement is accentuated in the light of digitalisation. The last three decades or so have witnessed the dawn of big data (Bharadwaj & Noble, 2015; Wedel & Kannan, 2016). Advanced technology enables customers to access information and communicate with other consumers and firms without time, location or financial constraints (Hoyer, Chandy, Dorotic, Krafft, & Singh, 2010). Customers feel empowered and have a greater desire for engaging through exchanging and collaborating with companies (Ernst, Hoyer, Krafft, & Soll, 2010). UGC are created and exchanged in a more rapid, transparent and frequent manner (Hennig-Thurau et al., 2010). Individual opinions can be directly delivered to others around the world via the Internet (Hennig-Thurau et al., 2004). The fast pace of new data generation in the era of the Internet means that firms must possess the strategic flexibility to not only capture customers' data but also to utilise it to enhance their service delivery (Bharadwaj & Noble, 2015). Accordingly, identifying and capturing value inherent in UGC in a timely manner has become a strategic imperative for many firms (Rust, Moorman, & Bhalla, 2010).

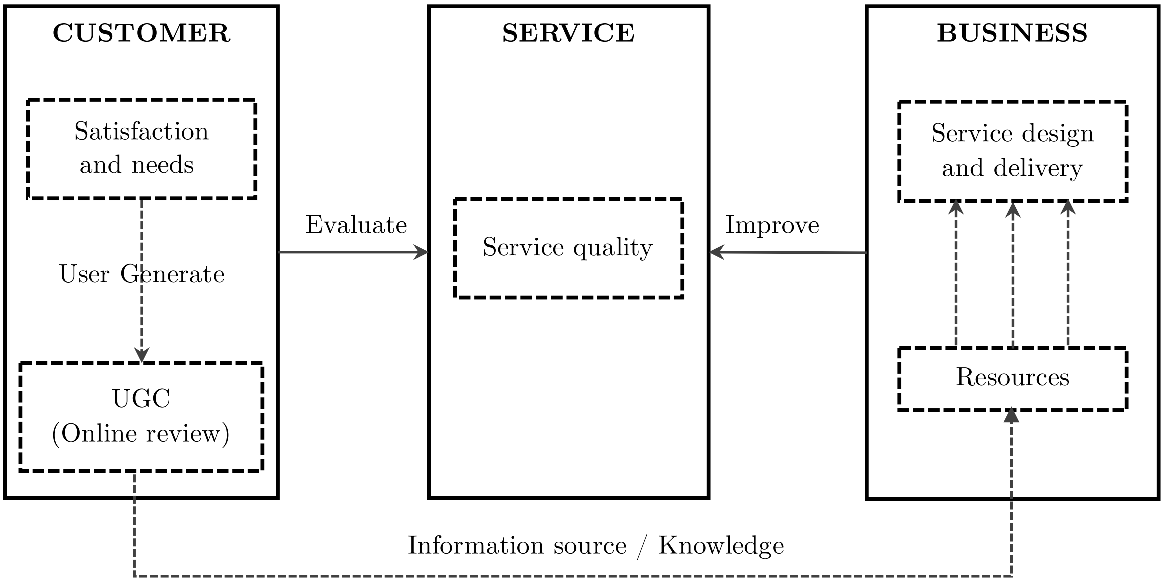


Figure 3-1. Conceptual map of using online reviews for service improvement

The greatest value of UGC in service improvement lies in the mediation between customer needs and firms' strategic intent (Cook, Bowen, Chase, Dasu, Stewart, & Tansik, 2002; Herrmann, Huber, & Braunstein, 2000). As a source of information, UGC is useful for understanding customers' behaviours and opinions (Kannan & Li, 2017), identifying unmet needs and pinpointing viable solutions (Berry, Shankar, Parish, Cadwallader, & Dotzel, 2006; Griffin & Hauser, 1993). The insights from user-generated data provide valuable knowledge for firms to design and improve services which are more in tune with customers' requirements (Bharadwaj et al., 2012; Cook et al., 2002; Kristensson, Gustafsson, & Archer, 2004; Prahalad & Ramaswamy, 2004). In this sense, a proactive strategy of listening to customers' online discussions is instrumental in service research and practice. Mining UGC is helpful for service design and improvement through revealing collective opinions of consumers, which enables managers' decision-making based on better understanding of the market. The awareness and anticipation of what customers value and the changes in customers' desires brings competitive advantage by driving customer satisfaction and loyalty (Flint, Blocker, & Boutin Jr, 2011). Based on the above analysis, it can be deduced that customers' comments can provide the fundamental pillar for firms to improve their service delivery system, as demonstrated in Figure 3-1.

3.3 Methodology

To illustrate the contention of mining UGC for service design and improvement, an analysis is conducted using customers' reviews on London hotels. As the interest lies in exploring customers' opinions from their written feedback, text analysis is performed on the unstructured data sources. The proposed methodology is illustrated in Figure 3-2. First, online customer reviews are collected from the website and the data is then cleaned in the MySQL database to remove duplicates, non-English reviews and special characters. On the basis of pre-processing the textual data, topic models are constructred to discover abstract 'topics' hidden in the large corpus, using the package MALLET (i.e., Machine Learning for Language Toolkit). Then the topics are interepreted to identify quality dimensions by collating service quality literature, and a decision model is developed prioritising the needs in service design and improvement.

3.3.1 Data

The raw data set containing over 810,000 online customer reviews of London hotels over 15 years is used in this study (as described in Section 1.5.3 in Chapter 1). Although these reviews reflect certain characteristics of hotels in one destination, reviewers are from all around the world and have diverse expectations and perceptions of hotel services given their different cultural and demographical background. This compensates for the geographic constraint in the data sampling.

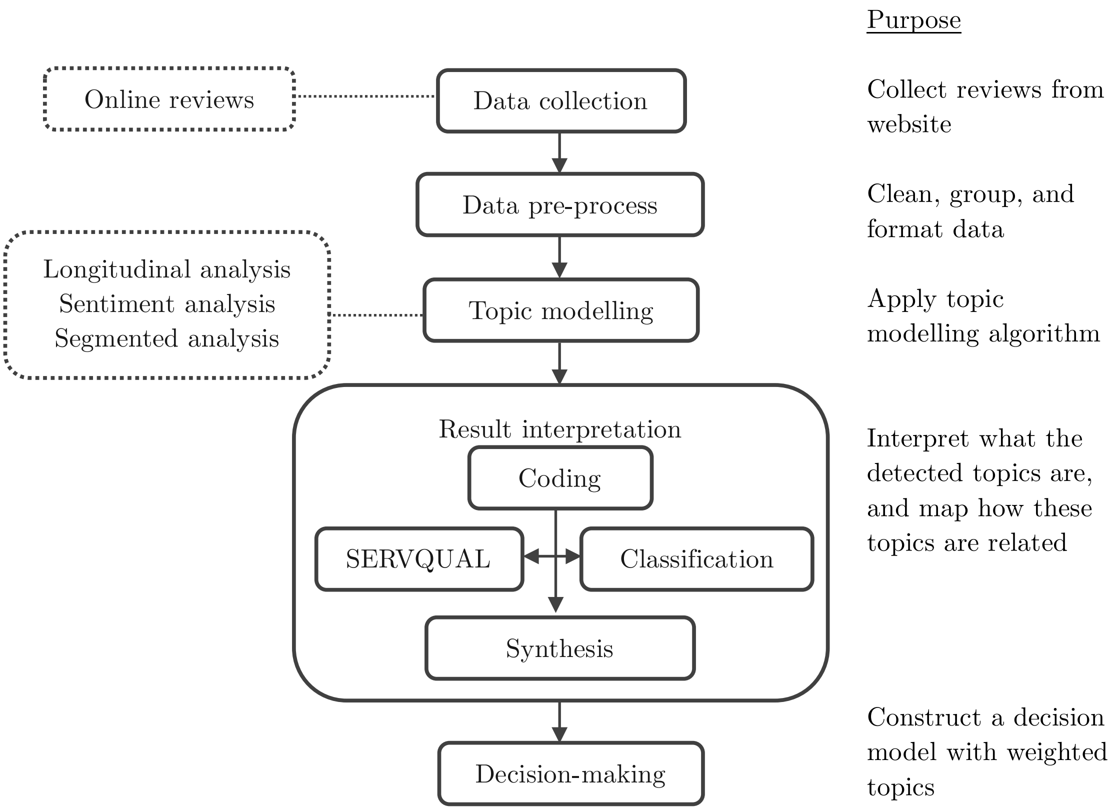


Figure 3-2. Illustration of the proposed methods

To ensure the cleanliness of the data, the entire data set is screened and duplicate records are removed. Reviews that are not written in English are also removed. This is accomplished through detecting languages with language indicators, such as the alphabet, special symbols and typical words unique to the language. About 180,000 non-English written reviews are successfully detected and deleted from the data list with careful inspection. This leads to a data set consisting of 632,076 reviews for the textual content and structure analysis. Next the reviews are tokenised by splitting the full text into a sequence of words that describe the content of the document (Feldman & Sanger, 2007). The choice of tokenisers such as delimiters, non-letters and specified characters, depends on languages and purpose of text mining. In this study, documents are divided using regular expressions, which results in a collection of tokens containing one or more characters. Tokens are further de-capitalised. Moreover, stop words which are functional words that do not carry any meaning are removed (Zhan et al., 2009). All these processes help reduce document complexity and noisy information in the subsequent topic modelling. The processed data of online customer reviews is from 1,028 London hotels covering a 15-year period from December 2001 to March 2016. Table 3-1 describes the data set.

Table 3-1. Data description

Full data set			
Number of hotels		1,028	
Average hotel class star		3.77	
Number of reviews		632,076	
Total number of tokens		27,986,649	
Max number of tokens in a review		1,836	
Average number of tokens in a review		44.28	
Average customer rating		4.01	
Data subset—Time			
Group	*Year*	*Number of reviews*	*Optimal topic number*
1	2001–2006	18,577	17
2	2007	10,575	18
3	2008	11,761	17
4	2009	19,027	17
5	2010	25,830	22
6	2011	45,283	23
7	2012	76,503	27
8	2013	105,514	31
9	2014	128,879	31
10	2015–2016	190,127	37
Data subset—Customer group			
Customer Group	*Rating*	*Number of reviews*	*Optimal topic number*
Satisfied	5	272,370	40
Neutral	3–4	286,359	42
Unsatisfied	1–2	73,347	29
Data subset—Market segment			
Market Segment	*Star Ranking*	*Number of reviews*	*Optimal topic number*
Upscale	5	112,600	35
Midrange	4	301,924	50
Midrange	3	176,343	40
Budget	0–2	41,209	24

Note: 0-star represents that hotels are unrated.

3.3.2 Topic modelling

Topic modelling is “a suite of algorithms that aim to discover and annotate large archives of documents with thematic information” (Blei, 2012, p. 77). It helps detect main themes pervasive in the unstructured text, uncovers the underlying semantic structure of a collection of documents and then organises a large corpus based on the identified topics (Blei, 2012). Natural language processing (NLP) and machine learning are involved in the construction process of the statistical model. Such machine learning is normally unsupervised, where no pre-defined categorisation or annotations are required, and topics emerge from the original text by applying a hierarchical Bayesian analysis (Blei & Lafferty, 2009).

Of all topic models, LDA is the simplest yet most efficient one that often serves as a springboard for topic modelling (Blei & Lafferty, 2009). LDA is a generative probabilistic model that builds on the tf–idf scheme (Salton & McGill, 1986), latent semantic indexing (LSI) (Deerwester, Dumais, Furnas, Landauer, & Harshman, 1990) and probabilistic LSI (Hofmann, 1999). The basic idea behind LDA is a three-level Bayesian model, where a document in a text collection presents various topics, and each topic is comprised of a fixed vocabulary based on probabilistic distribution (Blei, Ng, & Jordan, 2003). The LDA model aims to discover various topics that the documents represent and the relative importance of each topic in a document. Apart from the observable words in the document, the latent topic structure, including topics, topic distributions in a document and the topic assignment to each word in each document, are hidden in the corpus. To uncover this hidden structure, the LDA model draws on an inference algorithm which relies on the joint probability distribution over the visible variables (i.e., words in the document) and hidden variables (i.e., latent topic structure) (Blei, 2012). Essentially, the topics generated are clusters of similar words based on the probability of words’ occurrence, showing a repeating pattern of co-occurring terms in a corpus.

An LDA model is developed to identify the topics underlying online customer review text. To that end, a computing programme, MALLET, is adopted to implement the LDA algorithm and construct a topic model using the collected data. MALLET is a Java-based package for machine learning on text data, which can conduct statistical NLP, topic modelling, document classification, and more (McCallum, 2002). Its topic model package implements Gibbs sampling in an extremely fast and highly scalable way.

The exploration starts by looking at the whole sample of customers' reviews in general and identifying broad topics revealed in these textual documents. On top of the overview, topics varying with time, customer groups, and market segments are examined. The topic models are trained on each data subset separately to capture topic structures in each subset. To find a 'best fit' for the topic number, the perplexity of a held-out test set is calculated to evaluate the trained model (Blei et al., 2003). The algorithm is run several times with different pairs of hyperparameters, a varying number of topics and sampling iterations to optimise the results (Graham, Weingart, & Milligan, 2002). The topic modelling results indicate the detected topics and composition of topics in documents.

Interpreting the output of topic models first requires each topic to be defined. The topic extracted is a collection of co-occurring words. All topics are carefully read by the researcher and each topic is summarised with one term based on the key words listed on that topic. Topics are validated and compared to determinants of service quality discovered in current literature, particularly in the hotel sector. Moreover, the SERVQUAL construct is applied to group and frame the quality dimensions detected from online review text. The relative importance of each dimension is examined by calculating the weights. These findings are then used to develop a model prioritising the needs in service improvement decision-making.

3.4 Findings

3.4.1 General topics in customer reviews

3.4.1.1 Determining the number of topics

Determining an optimal number of topics is critical in topic modelling as it has a significant effect on the model output. A common approach used in existing research is to calculate the log-likelihood (posterior probability) by estimating the probability of unseen held-out documents given a training data set (Wallach, Murray, Salakhutdinov, & Mimno, 2009). Higher likelihood on a held-out test data set implies a better model. In Blei et al. (2003), the perplexity of a held-out test set is used as a measure in the statistical topic model evaluation. Perplexity, conventionally applied in language modelling, tells how 'perplexed' a probability model is in predicting new data. It is "equivalent to the inverse of the geometric mean per-work likelihood" and "is monotonically decreasing in the likelihood of the test data" (Blei et al., 2003, p. 1008).

A good model with better generalisation performance should have a lower perplexity score. The perplexity for a test set with M documents is:

$$perplexity(D_{test}) = exp\left\{ -\frac{\sum_{d=1}^{M} log\, p(w_d)}{\sum_{d=1}^{M} N_d} \right\}$$

The topic model is conditioned on three parameters, which are the Dirichlet hyperparameters α and β, and the number of topics K. The perplexity scores are calculated to determine the value of the key parameters. Following Blei et al. (2003), the full data set is split into a training set (90%) and a testing set (10%). First, a grid search of α and β is done with a fixed value of K at 50. Given the assumptions of $\alpha \propto 1/K$ and the default setting of $\alpha = 5/K$ and $\beta = .01$ applied in MALLET package, the parameters are determined from $\alpha = \{.002, .02, .1, .2\} = \{.1/K, 1/K, 5/K, 10/K\}$ and $\beta = \{.01, .1, 1\}$. Table 3-2 presents the perplexity scores of test data set with 12 pairs of hyperparameters. The best combination is $\alpha = 5/K$ and $\beta = .1$, which reaches minimum perplexity.

Table 3-2. Perplexity with varying α and β

A	α = A/K	β	Perplexity (training)	Perplexity (testing)
.1	.002	.01	2474.39	24372.91
.1	.002	.1	2597.57	21535.28
.1	.002	1	3174.12	23211.47
1	.02	.01	2186.01	22977.07
1	.02	.1	2297.09	21133.19
1	.02	1	3004.74	23813.15
5	.1	.01	2055.94	22686.61
5	.1	.1	2146.83	20652.54
5	.1	1	3288.56	26280.77
10	.2	.01	2058.59	22275.33
10	.2	.1	2140.18	20769.34
10	.2	1	3655.79	27953.47

Note: K = 50

After determining the best pair of hyperparameters for fixed K, the optimal K is searched for with $\alpha = 5/K$ and $\beta = .1$. The changing value of K and associated perplexity are listed in Table 3-3. When K is set to 70, the perplexity in test data reaches the lowest. The perplexity, however, increases with bigger K due to the overfitting problem.

Table 3-3. Perplexity with varying K

K	Perplexity (training)	Perplexity (testing)
2	2469.85	22840.96
5	2337.42	21925.39
10	2251.49	21771.51
20	2186.78	21313.69
30	2158.50	21198.43
40	2143.49	20897.22
50	2148.23	20866.13
60	2157.12	20777.69
65	2153.95	20732.95
69	2165.97	20699.82
70	2164.69	20641.65
71	2164.44	20694.90
75	2169.24	20736.45
80	2181.07	20745.88
90	2196.12	20814.81

Note: $\alpha = 5/K$ and $\beta = .1$

In addition to computational evaluation of the models, it should be taken into consideration whether the outcome of the topic model meets the expectation given the research goal. Many scholarly works (e.g., Wallach et al., 2009; Grant & Cordy, 2010) suggest that in specifying topic number there is no fixed threshold that can be applied to all situations and all data sets. The decision on this parameter also largely relies on human judgement about the level of granularity in the discovered topics. As explained in Barua, Thomas, and Hassan (2014, p. 626), a larger number of topics will produce “finer-grained, more detailed topics” whereas a smaller number of topics will produce “coarser-grained, more general topics”. The Dirichlet hyperparameters α and β can also affect the results of topic structure (Griffiths & Steyvers, 2004). The balancing point depends on how researchers would control the desired outcome. In general, some level of granularity in the topics is wanted and too coarsely trained models with superfluous details may not be ideal. The aim is to strike a balance with an expectation of a medium level of granularity where the topics are not overly sparse while remaining distinct from each other.

For the above reasons, K is set to 70, which provides desired characterisation. When training the topic model, the function of hyperparameter optimisation is turned on in MALLET and set the optimisation interval at every 20 iterations to get a better fit to data by allowing different levels of prominence among topics. Such parameters setting is expected to return a relatively finer-grained model. In addition, the number of key words included for each topic does not affect the model estimation but the process of making good sense of what each topic means. Key words are listed in an order from the highest weighted to the least. Fewer words may make it difficult to explain the theme while too many words may also be challenging to summarise. Therefore, 10 words for each topic are presented in order to help the interpretation of themes in the detected topics.

3.4.1.2 General topics

The initial modelling results from 632,076 reviews are grouped into 70 topics sorted in descending order according to their weight. A higher weight indicates the associated topic has been more frequently mentioned by the reviewers. Figure 3-3 shows the top ten topics using word clouds. Each word cloud presents one topic, represented by the topic ten key words for that topic with the size of words indicating the frequency of these key words. An important part of topic modelling is the interpretation and summarisation of topic themes. The labelling of topics follows the process illustrated in Guo et al. (2017). Given each topic is represented by a list of keywords, the starting point is to identify a logical connection among the most frequent words for a topic. A candidate topic name is determined and further tested by examining if the logical connection holds for the rest of words in the list. After merging topics that represent similar meanings, in total, 38 specific attributes of hotel service mentioned in the sample of customer reviews are discovered (Figure 3-4).

Figure 3-3. Top 10 topics from the reviews

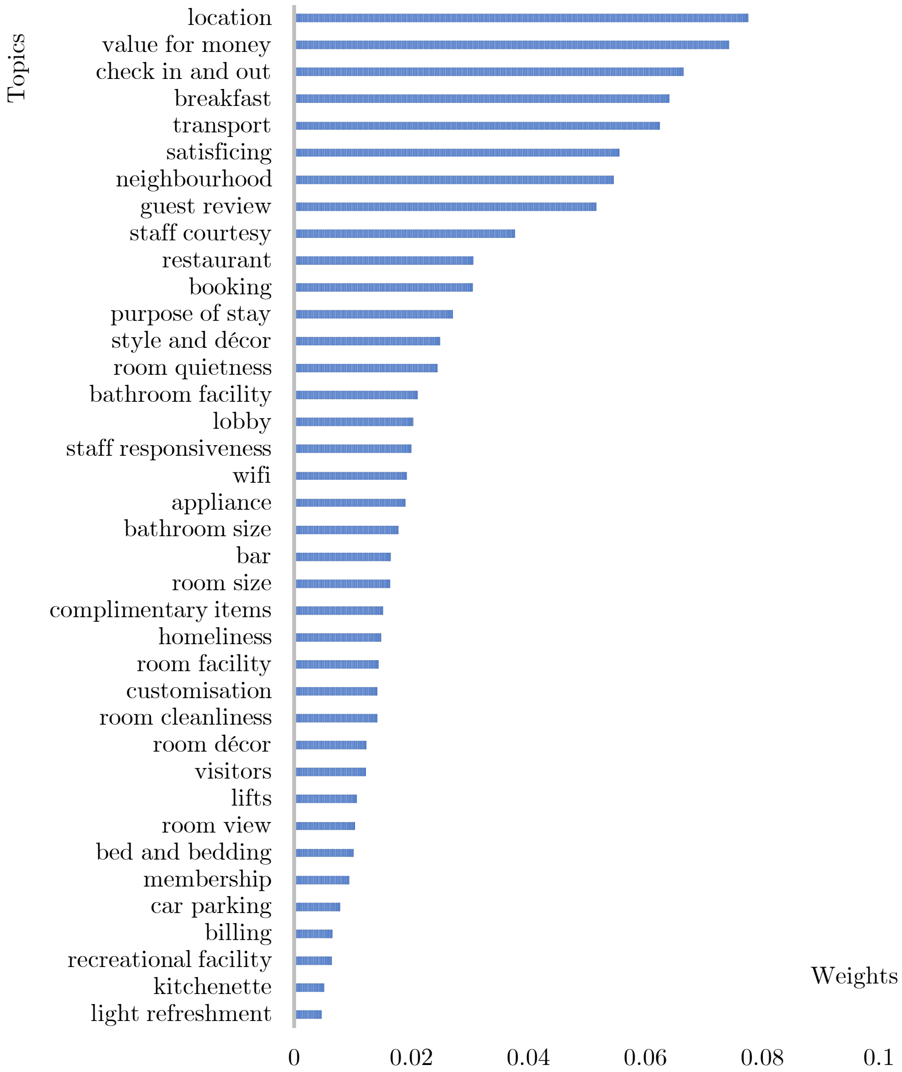


Figure 3-4. Important hotel service quality attributes from the reviews

Among the discovered attributes, location of hotel ranks the top in the review sample. Guests often describe where the hotel is situated, and comment on the ease of access to tourist attractions and city centre. Related to the locational importance, transportation connectivity (e.g., tube, bus, train, walking, taxi, etc.) and surrounding offering (e.g., car parking, restaurant, bars, shops, and grocery stores in the neighbourhood) are also important factors considered in customers' experience during their stay in a hotel. Furthermore, other top-ranked topics include 'value for money' and 'satisficing'. Customers evaluate the service by comparing their perception of experience to expectation. The formation of expectation is not only influenced by service standards indicated by the provider but also by opinions in previous guests' reviews. Review messages explicitly state whether the service quality has met their expectation and is worthy of recommendation. The satisficing is also disclosed in terms of the reasonability of prices, which reflects customers' appraisal of maximum benefits obtained from the services.

Another group of hotel service attributes concerns the people side. Promptness and accuracy in providing service are important to customers, particular on customers' arrival and departure. Customers comment on their experience regarding checking in and checking out, including efficiency, luggage handling, front desk operation hours, checking reservation and billing. In addition to the operation process, customers value the courtesy and helpfulness of staff in the direct interaction with the service provider. Professionalism, warm welcome and willingness to offer help are what customers expect from members of the front office. Besides, individualised care and a homely atmosphere make customers feel special and the experience memorable. Customers are willing to share their purpose of stay (e.g., leisure, business trip), the service they receive as a treat for a special event (e.g., room upgrade, room service) and the privileges they receive as membership holders of a hotel (e.g., room upgrade, lounge access). The remaining attributes show specific aspects of explicit service and tangibility associated with the hotel service, including guestroom (e.g., room size, cleanliness, view, facility, décor and complimentary items), dining service (e.g., restaurant environment, food), public area and service (e.g., style and décor, Internet connection, lifts, lobby, recreational facilities).

3.4.1.3 Comparison with existing literature

Extensive market research has been undertaken to explore the determinants of customer satisfaction, which is positively associated with service quality (Sureshchandar, Rajendran, & Anantharaman, 2002). To validate the topic model results, the findings from topic model analysis of review data are compared to the existing literature on attributes of service quality and determinants of customer satisfaction, especially in the hotel industry (see Table 3-4). As suggested by Tirunillai and Tellis (2014), the degree of overlap between quality attributes discovered from topic modelling results and prior studies is measured by the Jaccard coefficient (Jaccard coefficient = $N(Dim_{TM} \cap Dim_{LR})/N(Dim_{TM} \cup Dim_{LR})$). Dim_{TM} and Dim_{LR} respectively represent the set of quality attributes derived from the sampled customer reviews and the relevant studies. A higher value of the measurement illustrates a higher degree of overlap between findings in literature and results from the topic analysis.

Table 3-4. A comparison of service quality attributes between topic analysis and selected literature

Quality Attributes	Topic model	Prior studies	Quality Attributes	Topic model	Prior studies
Appliance	✓	✓	Membership	✓	×
Bar	✓	✓	Neighbourhood	✓	✓
Bathroom size	✓	✓	Purpose of stay	✓	✓
Bathroom facility	✓	✓	Recreational facility	✓	✓
Bed and bedding	✓	✓	Restaurant	✓	✓
Billing	✓	✓	Room cleanliness	✓	✓
Booking	✓	✓	Room décor	✓	✓
Breakfast	✓	✓	Room quietness	✓	✓
Car parking	✓	✓	Room service	×	✓
Check in and out	✓	✓	Room size	✓	✓
Complimentary items	✓	✓	Room facility	✓	✓
Customisation	✓	✓	Room view	✓	×
Guest review	✓	×	Safety and security	×	✓
Homeliness	✓	×	Satisficing	✓	✓
Hotel and tour guide	×	✓	Staff courtesy	✓	✓
Housekeeping	×	✓	Staff responsiveness	✓	✓
Kitchenette	✓	✓	Style and décor	✓	✓
Lifts	✓	✓	Transport	✓	✓
Light refreshment	✓	✓	Value for money	✓	✓
Lobby	✓	✓	Visitors	✓	✓
Location	✓	✓	Wi-Fi	✓	✓

Notes: Prior studies used in the comparison analysis include Antony, Antony, & Ghosh (2004); Berry, Zeithaml, & Parasuranman (1985); Cadotte & Turgeon (1988); Chu & Choi (2000); Guo et al. (2017); Kandampully & Suhartanto (2000); Li, Ye, & Law (2013); Magnini, Crotts, & Zehrer (2011); Min & Min (1997); Min, Min, & Chung (2002); Mohsin & Lockyer (2010); Ramanathan (2012); Saleh & Ryan (1991); Slevitch & Oh (2010); Zhou et al. (2014). Some attributes' names are not exactly the names given in individual studies. The terms developed at the topic interpretation stage are used. In addition, in some cases, the attributes discovered and named in the topic model results are more detailed than the ones mentioned in prior studies, and vice versa. The terms derived from the LDA analysis are used, and attributes covering details in both sets of results are matched. ✓ stands for included, × stands for not included. Jaccard coefficient is .786.

The Jaccard coefficient is .786. This implies that a few attributes derived from the topic modelling analysis are new compared to prior studies. Five attributes, namely Guest review, Homeliness, Room view, Membership and Purpose of stay are unique to the topic modelling analysis of the hotel review data set. Meanwhile, several attributes mentioned in the existing studies are not discovered in the review sample, e.g., safety and security issues, room service, hotel and tour guide, and housekeeping service. There is no match when relevant words such as security, safety, housekeeping, laundry and guidebook are searched for among the entire key words list containing the top 20 words for each topic. However, these terms appear in the dictionary generated from the entire corpus. It implies that these words carry little weight in the corpus, leading to a lower rank in the top key word lists of detected topics. As illustrated in the previous section, interpretations of topic themes are based on the first few top words that appear most frequently on that topic. This might be the reason why these attributes are not discovered in the automatic text analysis.

3.4.1.4 Evaluation model based on customer-perceived service quality

The topic modelling analysis outlines the hotel service quality attributes that are of greatest concern in online customer reviews. The detected attributes are further grouped according to the RATER factors in the SERVQUAL scale to measure their relevant importance in consumers' perception of service quality. SERVQUAL instruments, developed and refined by Parasuraman, Zeithaml, and Berry (1985, 1988, 1991, 1994), is a widely applied measure in service quality research and industry (Ladhari, 2009). Recognising that consumers' perception of service quality depends on their experience with the service provider and the disagreement between service perforamance and their expectation, the SERVQUAL model offers a scale containing 22 measurable dimensions in five facets, denoted as RATER (*Reliability, Assurance, Tangible, Empathy, Responsiveness*). Although there is a debate about the measurement (e.g., Cronin Jr & Taylor, 1992), the importance and significance of the SERVQUAL model are well noted in the literature (Seth, Deshmukh, & Vrat, 2005). In the hotel setting, some quality dimensions differ from the five factors described in the original SERVQUAL model, and many researchers have modified the scale to fit in the sector (as summarised in Akbaba, 2006). The intention of applying the SERVQUAL scale is to frame the topic modelling results and systematically examine the significance of each attribute in customers' online word-of-mouth communications.

As shown in Table 3-5, *tangible* aspects of a hotel service have a dominant influence on consumers' perception of service quality. It mainly involves the quality of physical appearance of the guestroom, food and dining environment, public areas and external area. Of these four aspects, external area is more important to consumers as location of hotel is critical to guests' access to transportation, tourist attractions and activity in the nearby area. Another principal determinant is the physical quality of guestroom, which is the core component of the hotel service package. Room quietness and cleanliness, decoration and facilities (e.g., furniture, heating, lighting), spaciousness (e.g., room size, suite) and better view, provision of appliances (e.g., hair dryer, kettle, coffee machine) and complimentary items (e.g., beverages, toiletries) are frequently mentioned in consumers' online comments. Moreover, dining experience, especially breakfast, is another important factor that consumers consider when evaluating a hotel service. Common issues commented on in the online reviews include food quality and range of choices, service promptness, and dining environment. In addition, appearance (e.g., building décor, lobby, style and atmosphere) and service provided in public areas (e.g., wi-fi, business centre) are also important for impressing guests.

Table 3-5. Weighted dimensions of service quality using SERVQUAL concepts

Category	Weights*	Definition**	Dimensions
Tangibles	.580	Physical facilities, equipment, and appearance of personnel	Guestroom, Dining, Public area, External area
Reliability	.163	Ability to perform the promised service dependably and accurately	Value for money, Billing and booking, Standard
Assurance	.093	Knowledge and courtesy of employees and their ability to inspire trust and confidence	Staff courtesy, Satisficing
Responsiveness	.086	Willingness to help customers and provide prompt service	Staff responsiveness, Promptness of check in/out.
Empathy	.078	Caring, individualised attention the firm provides its customers	Customisation, Homeliness, Knowing customers

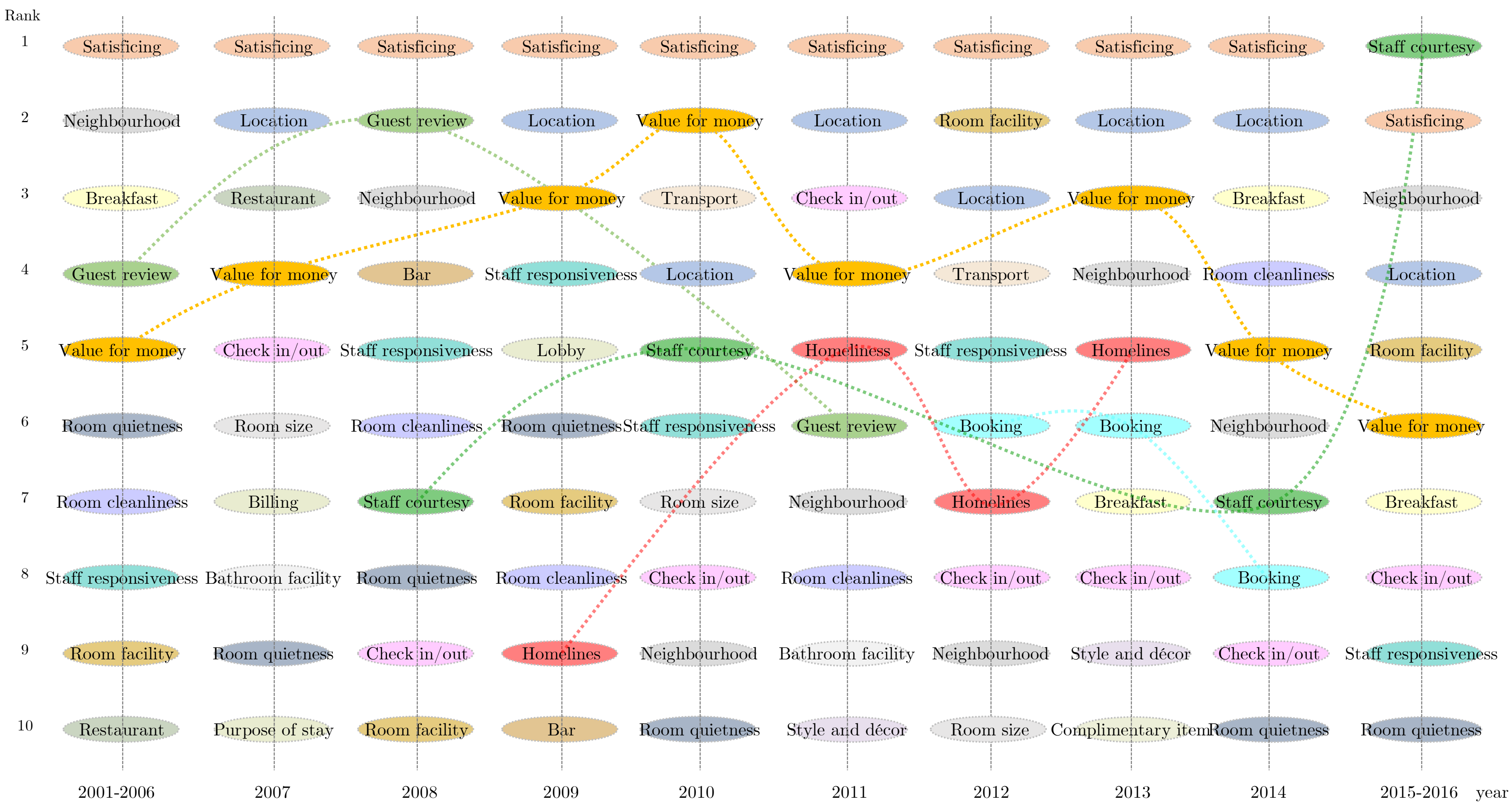
Notes: * This is measured by the sum of weights of all quality dimensions in each category.
** The RATER definitions are adopted from Parasuraman, Berry, & Zeithaml (1990, p. 35).

In contrast, the remaining factors are concerned with the intangible aspects of service, evaluating the behaviour and attitude of the service provider and its personnel. First, regarding *Reliability*, consumers would perceive the hotel as being reliable if it provides promised service to an expected standard and accurately performs the service (e.g., keeping accurate records of reservation and billing). The second important factor, *Assurance*, relates to the trustworthy relationship that is formed between the consumer and the service provider through their interactions. Staff courtesy is critical in this sense to create a friendly and comfortable climate to build trust and satisfy consumers. In terms of *Responsiveness*, promptness of service and helpfulness of staff during check-in and check-out, quick reaction to emergency (e.g., fire alarm), and handling complaints are critical to consumers' evaluation of service quality. In addition, in relation to *Empathy*, consumers are pleased to receive individualised and customised service. Knowing customers and offering personalised attentiveness, whilst ranking lowest in the five RATER factors, are becoming more vital to service quality.

3.4.2 Evolving topics over time

Give the data is from 2001 to 2016, it can be interesting to see how these topics changed over the 15-year time span. A closer examination of the shorter segment of time series enables us to track the shift in consumers' behaviour and preferences. Due to the relatively small amount of reviews in the early 2000s, reviews in the period from 2001 to 2006 are merged into one group and split the entire data set into ten groups (as shown in Table 3-1), and apply the same topic modelling approach outlined in the previous section to each group. Figure 3-5 maps the top ten themes discovered from yearly reviews, ranked based on their relative importance.

Overall satisficing is mentioned most frequently by consumers in all years' reviews. Customers also think highly of location, value for money, neighbourhood and check in/out of hotels. It is noticed that these dimensions and others such as room cleanliness, room quietness, breakfast, room facility and staff responsiveness are always mentioned by guests over the years, while some aspects appear in certain periods. For example, value for money seems more important to customers in the post-financial crisis period. Homeliness is particularly important around the year 2012, given that the London Olympic was held then and guests from all over the world might have emphasised on the homely atmosphere of hotels. Furthermore, customers in earlier years of the sample period indicated their observation of other customers' reviews; however, this attribute is less mentioned by customers in recent years. A possible explanation is that at the beginning of the development of online review platforms, customers value such reviews for their novelty and are likely to mention its usefulness when writing reviews. As review platforms have proliferated, an increased number of customers share opinions but less likely to highlight helpfulness of others' reviews give the fact that writing reviews is becoming a common practice. In addition, a few topics emerged in the later sample period. For instance, online booking became more important to customers from the year 2012. Reviews mentioning staff courtesy appeared more often after the year 2008 and this attribute ranked the most important in the recent two years.



Note: Topic frequency in the Figure—Satisficing-10, Check in and out-8, Location-8, Neighbourhood-8, Value for money-8, Room quietness-7, Staff responsiveness-6, Room cleanliness-5, Room facility-5, Breakfast-4, Homeliness-4, Staff courtesy-4, Booking-3; Guest review-3, Room size-3, Bar-2, Bathroom facility-2, Restaurant-2, Style and décor-2, Transport-2, Billing-1, Complimentary items-1, Lobby-1, Purpose of stay-1.

Figure 3-5. Top 10 themes in time-segmented reviews

Apart from examining the changes of top ten themes over the period, all the quality dimensions in each time segment are further grouped into the SERVQUAL construct to monitor how the significance of the RATER factors changed in these 15 years (see Figure 3-6). *Tangible* aspects of hotels appear most often in consumers' online reviews and *Assurance* of hotel service ranks the second. It is also observed that *Empathy* of hotel service has become more important to customers in recent five years compared to earlier years in the sample period. These results imply that the tangible sides remain the core of service offering, which are the basics for hotel services to be perceived by consumers as maintaining good standards. On top of this, consumers start to care more about the humanistic awareness of the service provider. This suggests that service providers should pay attention to consumers' changing mind-set of good quality and thereby improve service offerings accordingly.

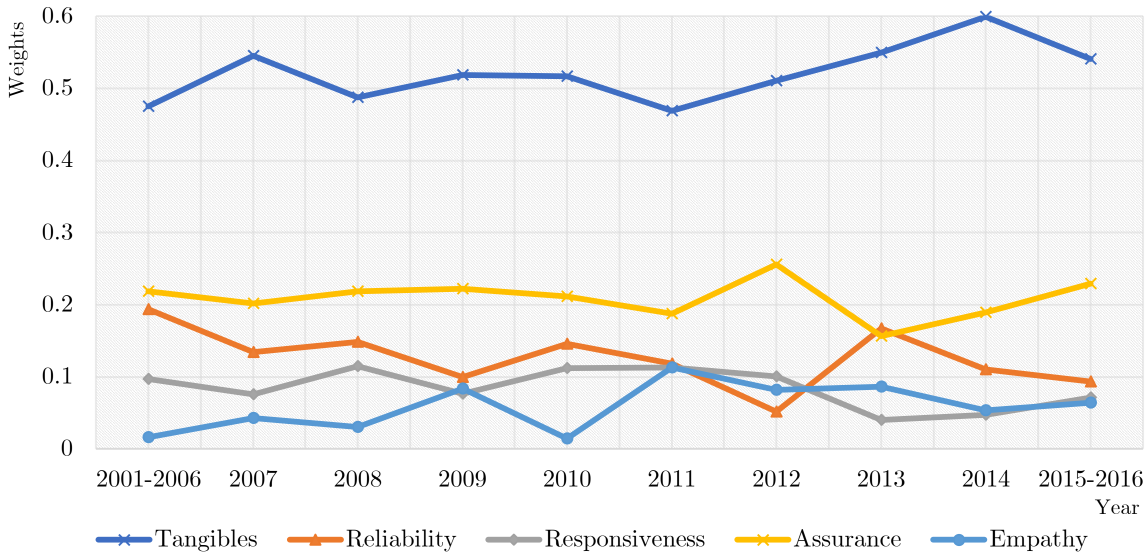


Figure 3-6. Yearly changes in the importance of RATER factors

3.4.3 Topics varying with customer sentiments

Considering that satisfied and unsatisfied customers may have different focuses when sharing their experience, data are grouped based on individual rating scores. The star ratings indicate the overall degree of customer satisfaction regarding received service quality. Ranging from 1 to 5, a higher score implies more positive attitude toward service quality, whilst a lower score shows more negative sentiment and dissatisfaction with service experience. Given that the average rating of the sample data is 4.19, three customer groups are defined, which are satisfied (i.e., higher rating than the average), neutral (i.e., average rating), and unsatisfied (i.e., lower rating than the average) customer groups, respectively with customer ratings at 5-score, 3- and 4-score, 1- and 2-score. A comparison of the importance of top ten quality attributes in the three customer groups is presented in Table 3-6. A particular interest is given to the attributes that lead to consumers' dissatisfaction with hotel service. There are a few areas such as value for money, check in and out, and room cleanliness that unsatisfied customers often complain about. Furthermore, according to the significance of RATER factors in the three groups (Figure 3-7), *Tangible*, *Reliability* and *Responsiveness* aspects of services seem to be the main causes for customers' dissatisfaction. Analysing customers' concerns expressed in the complaints may help service providers identify areas to deal with and improve quality of services.

Table 3-6. Top 10 themes in sentiment-segmented reviews

Rank	All reviews	Segmented Reviews		
		Higher-rating	**Average-rating**	**Lower-rating**
1	Location	Satisficing	Satisficing	Value for money
2	Value for money	Staff courtesy	Location	Check in and out
3	Check in and out	Location	Staff responsiveness	Room cleanliness
4	Breakfast	Bar	Neighbourhood	Satisficing
5	Transport	Breakfast	Value for money	Booking
6	Satisficing	Value for money	Breakfast	Bar
7	Neighbourhood	Check in and out	Lobby	Room size
8	Guest review	Room quietness	Room quietness	Staff courtesy
9	Staff courtesy	Customisation	Style and décor	Room quietness
10	Restaurant	Lobby	Check in and out	Restaurant

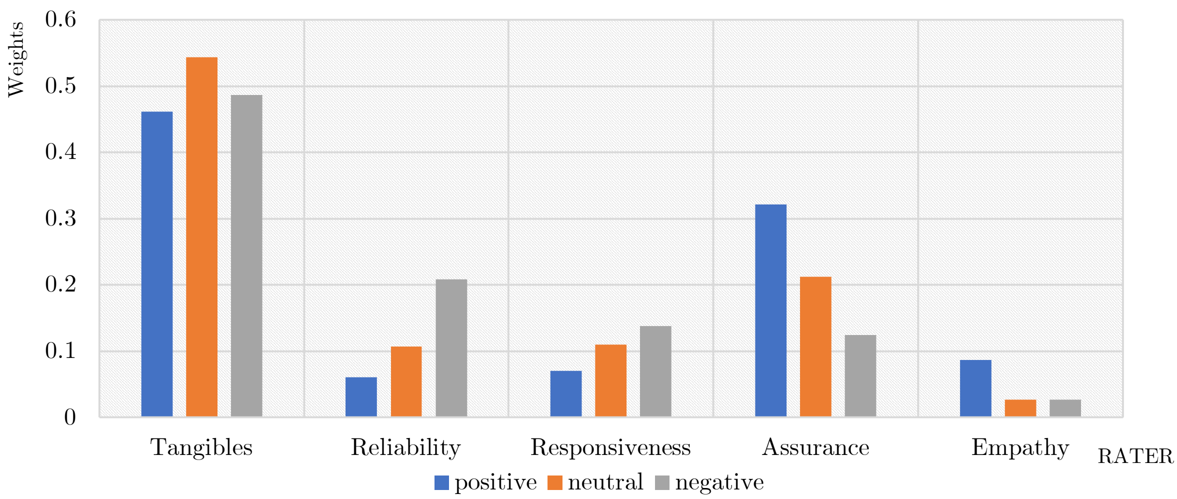


Figure 3-7. Importance of RATER factors in different customer groups

On the other hand, consumers may derive satisfaction from a higher level of *Assurance* and *Empathy* of service. The interaction between customers and service provider, as well as good manners and behaviour of staff that make customers feel cared for and welcome, are highly valued and appreciated by consumers. This is partly because that while customers may have an idea of the service standards of hotels through reading detailed information and peer consumers' reviews or previous stay experience, staffs' professionalism and courtesy may be an uncertain variable that could affect customers' feelings and perception. It suggests that more effort can be put into providing a customised and personalised service that is promised or even beyond consumers' expectations so as to improve customer satisfaction.

3.4.4 Topics varying with market segments

Given that customers going to different types of hotels may have different expectations, the same methods are applied to detect topics from reviews of hotels positioned in different market segments. Data subsets are created according to the hotel star ratings provided by the Automobile Association (AA). Although hotel classification systems are different across the world, it is to some extent an indicator of service quality, which impacts on customers' expectation about luxury, midrange and budget hotels (Knutson, Stevens, Patton, & Thompson, 1993). The star class is recognised from the lowest band of 1-star to the highest 5-star on the basis of professional inspection of hotel quality. In general, such ratings correspond to a classification of hotel type. Hence, the reviews are segmented into four categories according to the hotel star ranking, more specifically, Deluxe/Luxury (5-star), First Class/Superior (4-star), Mid-range/Standard (3-star) and Budget/ Economy hotels (0–2-star) (see a detailed description in Table 3-1).

Table 3-7 provides a snapshot of the top ten themes in each market segment benchmarked against the top ten topics discovered in the full data set. It reveals that some quality attributes are common across different market segments. To be specific, location, value for money, staff courtesy and check in/out have been frequently mentioned in the online reviews by customers in all market segments but their significance to the overall perception of service quality varies with hotel types. Meanwhile, there are some distinctive features in terms of customer concerns about service quality in different types of hotels. For instance, location is one of the top priorities for all hotels except the budget hotels as budget hotel customers are willing to compromise on the convenience for good value for money. In contrast, room quietness and transport are more often reffered to in the reviews for 2- and 3-star hotels. In addition to the top ten themes, it is also found that Wi-Fi has been regarded as an essential service for all hotels except the budget hotel while room view is only important for luxury hotel customers but not for others.

Table 3-7. Top 10 themes in market-segmented reviews

Rank	All reviews	Segmented reviews			
		2 Star	**3 Star**	**4 Star**	**5 Star**
1	Location	Value for money	Breakfast	Value for money	Satisficing
2	Value for money	Transport	Location	Location	Location
3	Check in/out	Check in/out	Value for money	Neighbourhood	Value for money
4	Breakfast	Staff courtesy	Neighbourhood	Satisficing	Staff courtesy
5	Transport	Room quietness	Satisficing	Style and décor	Style and décor
6	Satisficing	Room décor	Staff courtesy	Room size	Restaurant
7	Neighbourhood	Breakfast	Transport	Staff courtesy	Staff responsiveness
8	Guest review	Room size	Booking	Breakfast	Room cleanliness
9	Staff courtesy	Complimentary item	Room quietness	Check in/out	Check in/out
10	Restaurant	Billing	Check in/out	Transport	Booking

For hotels in different market segments, customers expect different standards of service and hence the centre of attention in their evaluation of service quality is different (see Figure 3-8). Across the four market segments, *tangible* aspects of a hotel service remain the most influential determinant of customers' perception of service quality, although its significance is slightly different across market segments—consumers who stay in the budget and mid-range hotels discuss the physical quality more than those staying in a luxury hotel. *Reliability* is the second important factor among all segments except the luxury hotel. In the high-end market, consumers care more about the *Responsiveness* and *Assurance* of hotel service, but staff courtesy and behaviour are less important in lower-rated star hotels. In addition, 4- and 5-star hotels experience higher expectations from consumers in relation to the *Empathy* side of service, compared to lower-end hotels. These findings imply that hotels should take care of both tangible and intangible aspects of their servicer offering, but prioritise dimensions of customer-perceived service quality based on the market and brand positioning.

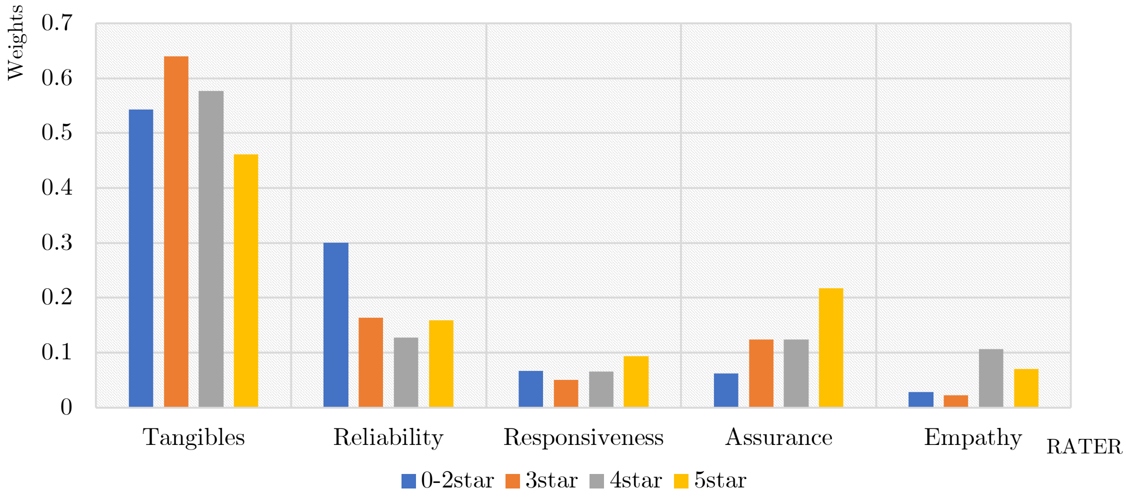


Figure 3-8. Importance of RATER factors in different market segments

3.5 Conclusions

In the light of growing recognition of the importance of harnessing big data in business decision-making (Vidgen, Shaw, & Grant, 2017), the current study explores how UGC can be leveraged to improve service quality. It is emphasised that businesses learning from UGC is a strategic imperative to improve service quality. This strategy is demonstrated by examining a large data set containing over 600,000 online customer reviews of London hotels using an LDA topic modelling approach. The topic modelling results capture the most predominant attributes perceived by customers who share their opinions and experiences on the online review platform. Some significant themes are discovered, for example, guest review, homeliness, room view and membership, emerge from the topic modelling analysis which have been overlooked by prior studies. On the other hand, it is observed that themes such as hotel and tour guide (Min & Min, 1997), housekeeping (Kandampully & Suhartanto, 2000), room service (Mohsin & Lockyer, 2010), and safety and security (Chu & Choi, 2000), that are highlighted in the existing literature, do not surface from the topic modelling analysis. This finding indicates some shift in consumer behaviour especially for the group of consumers being active online.

Moreover, the weighted dimensions of hotel service quality discovered from online customer reviews offer insights into consumers' perception and evaluation of quality. The analysis results reveal the primary concerns of hotel customers. For instance, location, value for money, and check-in/out are the most important service quality attributes for hotels and this applies to hotels across different market segments. Furthermore, it is found that physical-related tangible and people-related intangible attributes are both critical to customer-perceived service quality, although physical-related tangible aspects seem to carry higher weights. In addition, the segmented analysis demonstrates that consumers' concerns vary with time, customer satisfaction, and hotel types. Decisions on service improvement are therefore contingent upon the market segments where firms are positioned and upon the satisficing and problematic areas pinpointed by analysing customers' praises and complaints. These findings reinforce the need for firms to understand consumers' primary concerns and allocate resources accordingly, which is an important implication for the hotel and service sector.

Similar to many other previous studies, there are some limitations inherent in this type of research. For instance, the interpretation of topic modelling results is to some extent subjective. Future research may consider validating the findings from LDA analysis of users' voices by exploring whether the identified dimensions are consistent with real-world managerial activities. In addition, this research examines only the online reviews from London hotels. It would be valuable to incorporate data from other international cities to improve the robustness and applicability of these research findings. Finally, this research mainly focuses on applying this novel analytical approach to the hotel sector using online reviews. One valuable future direction is to extend the application to other industry sectors using other forms of UGC (e.g., social media data).

CHAPTER 4

STRATEGIES FOR RESPONDING TO CUSTOMER REVIEWS: ESTIMATING THE INTERACTIVE EFFECTS ON FUTURE RATINGS

4.1 Introduction

Over the last two decades, the growing number of Internet users and declining communication costs have stimulated the proliferation of user-generated content (UGC). UGC sites such as Twitter, Facebook and TripAdvisor create new social interactions among customers and also change the way firms interact with their customers (Hennig-Thurau et al., 2010; Sonnier et al., 2011). The online review and response system is a typical example. Customers can go to the website to share their experiences of the hospitality services offered by service providers, and managers have an option to add responses to these reviews. In practice, a firm's direct response on online review platforms has become a new channel to perform service recovery and to retain customers who have purchased, experienced and rated the service.

Meanwhile, both customer reviews and managerial responses are displayed on the site and read by a great number of online users, many of whom will be potential customers and reviewers. Given the public nature of management responses, the actions taken by firms may tell us more than the reviews themselves. To those customers who read online reviews, firms' responses and interactions with former customers may affect their inferences about firms' trustworthiness (Sparks, So, & Bradley, 2016; Wei et al., 2013) as well as their expectations and satisfaction (Gu & Ye, 2014). Responding to reviews conveys a signal to the public that firms highly value customers' feedback. A firm appears to be attentive and hospitable by responding appreciatively to customers' compliments and recognition of service quality. If an unfavourable review is displayed, an appropriate response such as an acknowledgement, explanation and offer to fix problems may help validate a firm's reliability and sense of responsibility. It is a demonstration of a firm's willingness to make the effort to improve product or service quality (Lee & Hu, 2005; Xie, Zhang, & Zhang, 2014).

From the above discussion, it is clear that a firm's involvement in the online review system may exert influence on future customers who recognise the online firm–customer interactions. What does this imply for the management of online customer reviews? A recent line of research has started to explore the significance of managerial responses in influencing potential customers. A few scholarly works have discovered that responding to online reviews can affect customers' evaluation of brand and perception of business trustworthiness (e.g., Lee & Song, 2010; Sparks et al., 2016; Wei et al., 2013). Empirical evidence on the association between managerial responses and business financial performance has been documented in a few studies (e.g., Xie et al., 2014; Xie, So, & Wang, 2017; Xie, Kwok, & Wang, 2017; Kim et al., 2015). Moreover, an emerging body of research shows interest in the potential influence of managerial responses on customer rating and review behaviour (e.g., Min, Lim, & Magnini, 2015; Proserpio & Zervas, 2017; Xie, Zhang, Zhang, Singh, & Lee, 2016).

Given the growing significance of online review systems in influencing customers' preferences and online engagement behaviours (e.g., Cascio et al., 2015; Goes et al., 2014; Lee et al., 2015; Sparks & Browning, 2011; Sridhar & Srinivasan, 2012; Vermeulen & Seegers, 2009), it is contended that identifying effective strategies to manage customer opinions on online platforms is an interesting issue. This is particularly important for firms in leveraging online resources to establish a positive reputation—the evaluation of a firm by customers in the online review and rating—which will help gain and sustain competitive advantages (Deephouse, 2000). Prior studies have found that responding to online customer reviews can be beneficial to online rating improvement (e.g., Gu & Ye, 2014; Proserpio & Zervas, 2017; Xie et al., 2016), but a more important question is how to respond once a business decides to provide managerial responses. Given that the implementation of managerial responses may yield positive outcomes, this study explores the mechanism that leads to such favourability in customer ratings. This contains a series of practical and unexplored questions, for example: Should all the reviews be responded to? Is responding speed important? Does the length of response matter? Should a standardised format be applied? What tone of voice should be used? And more importantly, what approaches are preferable given different characteristics of online reviews?

In this regard, current literature lacks a clear justification of the efficacy of managerial response strategies—the detailed nature of the responses in particular—in influencing customer ratings over time. This is an important research void considering that providing a response is not merely a one-off policy change, but a strategy that may have a continuous impact on customer ratings. This study attempts to examine the interactive effects of online managerial responses and customer reviews on future customer ratings. Specifically, employing signalling theory, expectation–disconfirmation paradigm, and consumer inference theory as the overarching theoretical anchor, this study develops and tests a model of managerial responses in terms of response ratio, speed, length, content standardisation and sentiment; its interplay with customer reviews; and their joint effects on subsequent customer ratings. To this end, this study analyses the data set containing over 800,000 online customer reviews and 360,000 attached managerial responses from a travel website. Using text analysis and regression approaches, it is found that an increase in future customer rating is associated with higher response ratio, prompt response and positive sentiment therein, and these effects are moderated by review valence.

The chapter considers managerial response as an ongoing strategy that may have a continuous impact in the long run. Different from a comparison before and after the policy change (e.g., Proserpio & Zervas, 2017; Wang & Chaudhry, 2018), this study examines how managerial responses to reviews in a period affect ratings in the subsequent period. It allows us to estimate the underlying dynamic effects in a time series, which deepens our understanding of its sustained impact on ratings after the initial response. Besides, instead of answering the question of whether to respond, this chapter attempts to delineate how to respond. Compared to prior studies that focus on the impact of the observability of managerial response on bystanders (e.g., Proserpio & Zervas, 2017; Wang & Chaudhry, 2018) or a single characteristic of response (e.g., Xie et al., 2016), the analysis investigates the detailed nature, including the textual characteristics, of managerial responses (i.e., response ratio, speed, length, standardisation, sentiment) and their possible effects on customer ratings. It extends the current line of research on this topic by simultaneously testing the significance of these characteristics and identifying the most important attributes that affect future ratings.

Moreover, management response is a multifaceted problem involving an interplay of the customer-generated and firm-generated content. It is recognised that the effects of managerial responses on ratings may vary with review metrics (i.e., review valence, volume, variability—see Babić Rosario, Sotgiu, De Valck, & Bijmolt, 2016; You, Vadakkepatt, & Joshi, 2015). Taking account of the interplay between review and response, their joint effects on subsequent ratings and how the efficacy of managerial responses is contingent upon the nature of reviews are tested. Deviated from a focus on the customer-complaining scenario (e.g., Min et al., 2015), this study offers a clearer picture of how to manage different types of reviews in certain circumstances. A more comprehensive analysis of the multifaceted issue is presented, which advances our understanding of the interdependence between online reviews and managerial responses. This contributes to services management specifically on the relationship between organisational actions and customer behaviour (Subramony & Pugh, 2015).

In addition, text-mining techniques are applied to quantify the unstructured textual data. The text-mining approaches enable us to obtain word counts, calculate response text similarity and extract sentimental orientation of managerial responses. These attributes reflect the communication styles of managerial responses, which have been largely overlooked in prior studies. Findings from the data-driven analysis (both structured and unstructured data) document new evidence for the efficacy of managerial responses in influencing online ratings based on a large set of field data. It offers important managerial implications for a firm's engagement in online review networks.

4.2 Conceptual Framework and Hypothesis Development

4.2.1 Theoretical background

A firm's engagement in social media is a nascent area of interest. There are a handful of studies in this emerging line of research that study the relationship between online managerial responses and customer ratings. Gu and Ye (2014) explore how online responses affect future satisfaction of returning customers who have made an online complaint. They find that responding to online complaints is highly effective in increasing satisfaction of complaining customers, while concurrently decreasing the satisfaction of those who complain but do not receive a response. Min et al. (2014) conduct an experiment to examine the impact of responses to negative reviews and suggest that an empathy statement with rephrasing the main complaint is more effective in improving potential customers' satisfaction. Xie et al. (2016) present evidence on the effectiveness of providing a management responses in increasing future customer ratings and volume of electronic word-of-mouth (eWOM). In addition, through comparing two online travel websites—one with and one without managerial responses—Proserpio and Zervas (2017) discover that both rating scores and volume of positive reviews are higher after firms start responding to online reviews. A trade-off is that although negative reviews become less frequent, their length and detail increase. Wang and Chaudhry (2018) also discuss the externalities of management responses on bystanders and emphasise the divergent effects of responses on review valence. It is revealed that responding to negative/positive reviews can yield a positive/negative impact on subsequent customer ratings, and responses tailored to reviews amplify such divergent impact.

This study concurs with the aforementioned research that managerial responses may exert significant influence over prospective reviewers' ratings who have observed both the user-generated reviews and firm-generated responses. The understanding of management response influence on future customer ratings is drawn from three perspectives. The first views management responses as a signal to the public, which has a direct influence on future ratings. Signalling theory emphasises that signallers deliberately convey positive and imperceptible organisational attributes to receivers (Connelly, Certo, Ireland, & Reutzel, 2011) from which signallers benefit by receivers acting on the signals (Bird & Smith, 2005). The action of responding may signal to the public that firms are attentive, caring and responsive, leading to more positive attitudes towards the company and thus the following ratings. Nevertheless, a possible side effect of the presence of responses, which signals firms' willingness to listen to customers' opinions, is that it may encourage more complaints to be voiced (Chevalier, Dover, & Mayzlin, 2017).

Second, given that responses are available on the Internet, consumers may form more realistic expectations through observing others' experience of consumption and interactions with firms. The underlying assumption is consistent with the expectation–disconfirmation paradigm of customer satisfaction (Anderson & Sullivan, 1993; Churchill Jr & Surprenant, 1982; Oliver, 1980). Responses can reveal latent qualities of products or services by delivering information about unobservable attributes, which is an additional source offered for individuals' consideration and assessment (Ye, Gu, & Chen, 2010). Through acknowledging the strengths or deficiencies revealed in the reviews, customers who observe the online conversations may adjust their expectations accordingly (Xie et al., 2016). In this regard, the level of satisfaction is possibly increased with a positive disconfirmation of beliefs. However, excessive positive reviews and positive responses may lead to higher customer expectations, which put greater pressure on the firms to satisfy the expectations.

In addition, the presence of multiple cues in managerial communications may affect new customers' inference about the organisation, according to the customer inference theory (Kardes, 1993; Kardes, Posavac, & Cronley, 2004; Kardes, Posavac, Cronley, & Herr, 2008; Sparks et al., 2016). Effective managerial responses may nudge negative impressions towards being positive. Responding in an appropriate manner shows firms are willing to maintain effective interactions with customers and view responses as an important part of management activities. This shapes customers' inference about firms' professionalism, trustworthiness and the quality of managerial communications, which further exert influences on their post-consumption rating behaviour.

4.2.2 Effects of managerial responses on ratings: Hypothesis development

According to the previous discussions, the managerial response may potentially have symbolic and informational significance. Therefore, this study identifies five important attributes associated with these functions, which are response ratio, speed, length, standardisation and sentiment of managerial responses (see Figure 4-1).

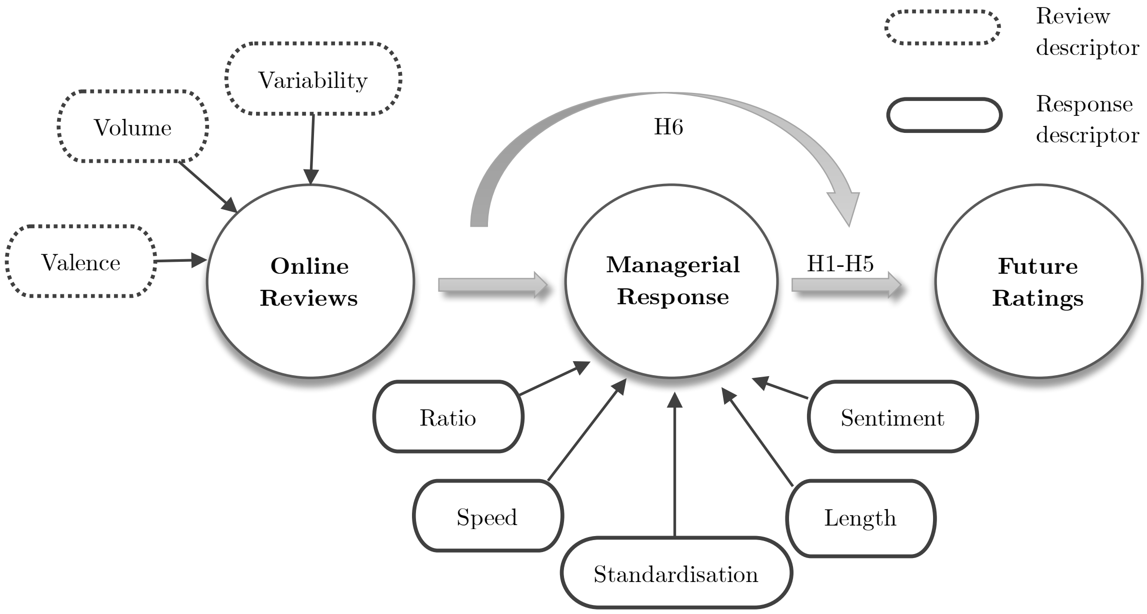


Figure 4-1. Research framework

4.2.2.1 Response ratio

Our first question concerns whether a higher volume of responses is preferable. When browsing the website, the number of responses displayed in the review section is the most visible element of online responses. It is the first signal delivered to review readers that a firm is responsive to customers' concerns. As discussed earlier, compared to no action, the existence of a response is likely to enhance ratings. A greater volume of response may be a positive signal to the public that a firm is responsive. It may imply that providing a response is an ongoing strategy rather than an ad hoc event and that the firm invests effort in reading and replying. Given that review volume varies across firms, the analysis uses response ratio (the percentage of reviews that are replied to) and hypothesises that:

Hypothesis 1: Response ratio is positively associated with subsequent customer ratings.

4.2.2.2 Response speed

The speed of responding measures the time intervals between a customer review post and the corresponding reply from a manager. The importance of timeliness has been widely documented in the service evaluation and recovery literature in the offline settings (e.g., Davidow, 2003; Wirtz & Mattila, 2004). In an online environment, proactive and timely responses are proven to be more effective in managing customer complaints (Homburg & Fürst, 2007; Tripp & Grégoire, 2011). Customers' inference about a firm's trustworthiness is likely to be influenced by the promptness of managerial responses (Sparks et al., 2016). A rapid response indicates organisational efficiency in dealing with customers' feedback, leading to a positive evaluation of a firm's controllability. The timeliness also affects the observability of responses, which is of critical importance to onlookers to notice the firm–customer conversations (Wang & Chaudhry, 2018). A delay in responding may lead to responses not being observed and therefore users are less likely to be affected by the responses (Proserpio & Zervas, 2017). Thus, it is hypothesised that:

Hypothesis 2: Response speed is positively associated with subsequent customer ratings.

4.2.2.3 Response length

The word count of a response text can be important. Research on eWOM has demonstrated that length of review message has direct or indirect effects on review readers' perception of communication quality (e.g., Baek et al., 2012; Mudambi & Schuff, 2010). Anchored in this line of thinking, it is assumed the length of response may make a difference to potential customers likewise. Longer responses cost firms more time and effort to write, which is likely to suggest that managers take customers' comments seriously. Positive customer perception and satisfaction may be derived from firms valuing the opportunity to communicate with engaging customers. Besides, analogous to that detailed reviews can reduce product or service quality uncertainty by revealing greater details (Mudambi & Schuff, 2010), longer responses may also increase such information diagnosticity (Menon, Raghubir, & Schwarz, 1995). Based on the uncertainty reduction theory (Berger & Calabrese, 1975), managers can provide more detailed information about the service, which would be otherwise not available to potential customers. Particularly in relation to those very specific comments from former customers, a more detailed explanation from the management helps customers develop realistic expectations of the commodity before purchasing or self-address some concerns at the post-consumption stage. Therefore, a longer response is potentially more informative and powerful in influencing customers. It is hypothesised that:

Hypothesis 3: Length of management responses is positively associated with subsequent customer ratings.

4.2.2.4 Response standardisation

Response standardisation is concerned with whether a response is tailored to each review. This study distinguishes two styles—standardised response and specific response. Responses written in a standardised way are generic and less relevant to the issues mentioned in a certain review. Future customers may interpret the highly standardised responses as being a lack of sincerity. Besides, repeated content adds little value to customers who rely on information cues presented in responses to make decisions. It is futile in terms of generating a positive impact on review readers. In contrast, a response that is specific to the related reviews seems more effective in improving ratings of prospective customers (Min et al., 2015). The textual content of a personalised response is more pertinent to particular issues raised in the reviews (Wei et al., 2013). It is an indicator of managerial effort being devoted to high-quality online communication, resulting in positive evaluation of the firm (Sparks et al., 2016; Van Noort &

Willemsen, 2012). In this way, responses play a more constructive role in positively influencing potential customers' inference and perception of service standard. Therefore, it is hypothesised:

Hypothesis 4: Standardisation of management responses is negatively associated with subsequent customer ratings.

4.2.2.5 Response sentiment

Sentiment displayed in a managerial response refers to the verbal (word) and more importantly the non-verbal (tone of voice) facets of the online communication. It reflects a firm's attitude towards vocal customers. According to the interactional justice theory (Bies & Moag, 1986), the interpersonal treatment is critical in determining individuals' reaction to an organisation. It highlights the significance of the manner in which customers' opinions are handled. The politeness and respect shown in the communications may be even more significant in influencing customer evaluation of firms. It is a relatively subtle aspect, but yet different strategies (e.g., empathetic or defensive) could have a divergent impact on customer evaluation (e.g., Lee & Cranage, 2014; Lee & Song, 2010; Min et al., 2015). A positive response puts customers' concerns first, for example, appreciating customers' praise, accepting responsibility for failure, taking corrective or preventive actions, and making promises to change in the near future. It is written in a courteous and businesslike way, leading to a positive impression of the firm's professionalism. On the contrary, a negative response presents no sincerity, which is manifest in the denial of accountability for unfavourable incidents, taking no action or paying lip-service towards failure, pointing an accusing finger at others, or speaking in an aggressive tone. Negative sentiment in the responses may result in adverse effects on future customers' satisfaction and ratings. It is reasonable to hypothesise that:

Hypothesis 5: Sentiment of management responses is positively associated with subsequent customer ratings.

4.2.2.6 Moderating effect

The development of responding strategies largely depends on the nature of word-of-mouth (i.e., valence, volume and variability of reviews—see Babić Rosario et al., 2016; You et al., 2015) and organisational factors (Lappas, Sabnis, & Valkanas, 2016; Lee, Xie, Besharat, & Tan, 2017;

Liu, Kim, & Pennington-Gray, 2015). In the positive review scenario, responses normally reinforce consumer's favourable statements and bolster a perfect image to the public. In the face of negative comments, potential customers may be more sensitive and pay more attention to the details. Inaction or inappropriate reaction (e.g., delay in responding, not specifically addressing the issues, defensive behaviour etc.) could cause greater damage to a firm's reputation by leaving negative reviews un-responded to or unchallenged. Thus the approaches used to respond to negative reviews and their effectiveness may differ from those when responding to positive ones. However, such potential discrepancy may be mitigated if customers also take account of the volume and variability of reviews. For example, when a hotel receives vast numbers of reviews, customers may excuse the firm for responding less in amount and slower in speed. If a higher degree of variability in ratings is presented, efficient responses may be required, from which customers draw inferences to eliminate uncertainty about the service quality. Accordingly, the effect of managerial responses on subsequent ratings is contingent upon the characteristics of customer reviews. It is hypothesised that:

Hypothesis 6: The influence of management responses on subsequent customer ratings is moderated by review characteristics (volume, valence and variability).

4.3 Methods

4.3.1 Data

Given the particular importance of firm–customer interactions for the service industry, this study examines the managerial response effects using data from the hospitality sector (as described in Section 1.5.3 of Chapter 1). The raw sample includes 813,287 customer reviews and 368,758 attached management responses of 1,063 hotels in London over 15 years. To clean the data, 26 hotels without customer reviews and duplicates in the data set are removed. Besides, reviews provided by the same customers on one hotel are also excluded given that the mechanism of management response influencing returning customers' satisfaction might be different. The final sample includes 723,680 customer reviews and 315,611 management responses (after removing non-English responses), with the oldest review being from December 2001 and the most recent from February 2016.

4.3.2 Text analysis

Text analysis is performed on response content to analyse the textual features of management response (i.e., length, standardisation and sentiment). This process quantifies the unstructured text into structured data for the subsequent statistical analysis. Data pre-processing is carried out in the MySQL database and the processed data is then imported into RapidMiner Studio, an industry-leading open-source platform for data mining, machine learning and predictive analytics (Hofmann & Klinkenberg, 2013), for text mining.

Data pre-processing. First of all, non-English responses are removed to ensure the consistency and accuracy of text analysis results. Many methods are available for language detection, but proprietary solutions are normally costly and some open-source solutions need complicated coding. By scanning the data set, some patterns in the response messages are found. In general, managers would take this chance to thank reviewers for taking the time to write the comments and apologise for any unpleasant experience. Specific words occur frequently in these common expressions, such as 'dear guest', 'thank you', 'grateful', 'appreciate', 'sorry', 'apologise', 'regards' and 'please'. Hence it might be possible to discern the language using our knowledge and common words in different languages. In addition, language indicators are also helpful for identification, for instance alphabet, special symbols and typical words that are the most frequent or unique for the language. Combining these two aspects, the screening approach is applied manually by running queries in MySQL with common expressions in different languages and strong language indicators. 19,334 responses written in non-English language are deleted from the database with careful inspection. Stop words and duplicates are removed, and all characters are decapitalised to reduce noise before importing the data into RapidMiner.

Document representation. In order to reduce document complexity and transform unstructured text data into a format available for statistical analysis, the full text of documents is split into a sequence of tokens describing the content of the document (Feldman & Sanger, 2007). There are several options for separating a document text, such as delimiters, non-letters, specified characters, named entities, regular expressions, and linguistic sentences or tokens. Non-letter characters (e.g., whitespace, punctuation) are used as splitting points, resulting in tokens at the word level. The number of tokens, representing the word count of each response, is recorded as a measure of the response length.

Text similarity. The standardisation of management response is examined by measuring the content similarity of responses provided by each individual hotel. A vector space model is used to determine the similarity between documents. Following the above document representation, tokens of a document are used to generate a vector, which numerically represents the document. The schema for vector creation chosen here is the tf–idf (term frequency–inverse document frequency). This weighting approach reflects how important a term is to a collection of documents by considering how frequent the term occurs in general. Then the cosine similarity in each pair of responses belonging to the same hotel is calculated. The similarity matrix tells how related two documents are by measuring the cosine of the angle between them on a multi-dimensional vector space with each term having an axis. The similarity score is bound between 0 and 1, where a greater value indicates a smaller angle between the documents. It implies the documents share more words in common, that is, a higher degree of standardisation in the textual content. Next, each response's similarity scores in relation to other responses from the same hotel are aggregated to get a mean value of content similarity. The overall average similarity score of all sampled responses is .078, with the lowest at .0009 and highest at 1.

Sentiment analysis. For the sentiment of responses, sentiment analysis is conducted to extract and classify the emotional orientation. First a training data set is created by randomly selecting 350 response examples and assign positivity or negativity to each response manually based on the classification scheme introduced in the hypothesis development section. The training data set is then imported into RapidMiner to construct a classification model using a linear Support Vector Machine (SVM) learner with a ten-fold cross-validation. The statistical performance of the SVM learner achieves 92.57% accuracy and the prediction precision achieves 72.92% and 95.70% for the two classes respectively. The model is then applied to all the vectored text and each response is labelled as positive or negative with a confidence score ranging from 0 to 1. The confidence score indicates the strength of prediction and in other words, the degree of the polarity. The confidence level of positive class is obtained as the sentiment score, with a higher score indicating that the response text has a more positive and supportive statement. The average sentiment level in the sample reaches .686, with the lowest at .075 and highest at .951 (see Table 4-1).

Table 4-1. Sentiment analysis: Model performance and results

SVM performance	True negative	True positive	Class precision
Pred. negative	35	13	72.92%
Pred. positive	13	289	95.70%
Class recall	72.92%	95.70%	
Prediction results	Confidence Score		
	Min	Max	Average
Positive Responses: 316,210	.075	.950	.686
Negative Responses: 26,073	.050	.926	.314

4.3.3 Variables and model specification

This study aggregates daily reviews and responses to monthly level (t) and organises all the data in a hotel–month panel. Table 4-2 presents a description of variables and summary statistics. The dependent variable is online ratings (*Rating*) measured as the average customer rating of hotel i in period t. Rating is modelled as a function of two groups of independent variables. The first group of explanatory variables includes descriptors of managerial response, which are the response ratio (*ResponseRatio*), the average response days (*ResponseDays*), the average length of response text (*ResponseLength*), the average degree of response content standardisation (*ResponseStandardisation*) and the average sentiment scores of responses (*ResponseSentiment*) of a hotel i in period t-1. The other group of explanatory variables consists of descriptors of customer reviews, including the average of customer ratings (*ReviewValence*), the number of customer reviews (*ReviewVolume*), and the standard deviation of customer ratings (*ReviewVariability*) of a hotel i in the previous t-1 periods.

Table 4-2. List of variables

Variables	**Description**	**M**	**SD**	**p50**	**Min**	**Max**
Dependent variable						
$Rating_{i,t}$	Average customer ratings (on a scale of 1–5) of hotel i in period t	3.550	1.074	3.870	1	5
Independent variables						
$ResponseRatio_{i,t\text{-}1}$	Ratio of the number of management responses to that of customer reviews of hotel i in period t-1	.221	.371	0	0	1
$ResponseDays_{i,t\text{-}1}$	Average number of days between management responses and the	39.95	214.3	6.667	0	4,037

	associated customer reviews of hotel i in period $t\text{-}1$					
$LnResponseDays_{i,t\text{-}1}$	The logarithm of $ResponseDays_{i,t\text{-}1}$	2.099	1.266	1.897	-.693	8.303
$ResponseLength_{i,t\text{-}1}$	Average number of words of management responses of hotel i in period $t\text{-}1$	90.96	56.34	80.62	4	2,061
$LnResponseLength_{i,t\text{-}1}$	The logarithm of $ResponseLength_{i,t\text{-}1}$	4.381	.504	4.390	1.386	7.631
$ResponseStandardisation_{i,t\text{-}1}$	Average score of similarity in the content of management responses of hotel i in period $t\text{-}1$	.087	.061	.073	0	1
$ResponseSentiment_{i,t\text{-}1}$	Average sentiment scores of management responses of hotel i in period $t\text{-}1$	.658	.095	.675	.101	.914
$ReviewValence_{i,t\text{-}1}$	Moving average of customer ratings of hotel i in the previous $t\text{-}1$ periods	3.501	.855	3.666	1	5
$ReviewVolume_{i,t\text{-}1}$	Aggregated number of customer reviews of hotel i in the previous $t\text{-}1$ periods	304.9	480.4	131	1	8,947
$LnReviewVolume_{i,t\text{-}1}$	The logarithm of $ReviewVolume_{i,t\text{-}1}$	4.747	1.568	4.875	0	9.099
$ReviewVariability_{i,t\text{-}1}$	Moving standard deviation of customer ratings of hotel i in the previous $t\text{-}1$ periods	.529	.280	.541	0	2.828
Control variables						
$Star_i$	Hotel star class evaluated by the third-party agency of hotel i	3.316	1.139	3	0	5
$LnSize_i$	Room number (logarithm) of hotel i	4.282	1.036	4.205	.693	7.396

Among these variables, *ResponseDays*, *ResponseLength*, and *ReviewVolume* are taken the logarithm values to achieve close-to-normal distributions. Given the fact that there is often a time lag between a managerial response and a review post and word-of-mouth effects are often cumulative and last for several weeks (Xie et al., 2014), lagged explanatory variables are used to examine the influence of management responses to the previous period's reviews on later customer ratings. Based on the Akaike information criterion (AIC) and Bayesian information criterion (BIC) statistics, one lag of response and review variables is considered. The analysis uses the time stamps of review posts and therefore the response variables demonstrate hotel *i*'s responses to time *t-1*'s reviews. With regard to the review variables, the average ratings, amount of reviews, and standard deviation of ratings of hotel *i* in the previous *t-1* periods are calculated, as the aggregated values are displayed on the site which may affect ratings in time *t*. Next, to rule out multicollinearity which may cause unreliable estimates of coefficients, the variance inflation factor (VIF) is examined and the results show that all VIF scores are less than 2 (see

Table 4-3), well below the common threshold. Accordingly, multicollinearity is not an issue in the model. Furthermore, the stationarity of all variables is checked with two Fisher-type unit-root tests (i.e., Dickey-Fuller test and Phillips-Perron test). The results provide sufficient evidence to reject the null of unit roots, suggesting that all variables are stationary at all panels.

Table 4-3. Collinearity diagnostics

Variable	VIF	SQRT VIF	Tolerance	R-Squared
$ResponseRatio_{i,t-1}$	1.130	1.060	.887	.113
$LnResponseDays_{i,t-1}$	1.170	1.080	.855	.145
$LnResponseLength_{i,t-1}$	1.070	1.030	.938	.062
$ResponseStandardisation_{i,t-1}$	1.120	1.060	.894	.106
$ResponseSentiment_{i,t-1}$	1.200	1.090	.835	.165
$ReviewValence_{i,t-1}$	1.510	1.230	.661	.339
$LnReviewVolume_{i,t-1}$	1.540	1.240	.651	.349
$ReviewVariability_{i,t-1}$	1.590	1.260	.630	.370
$LnSize_i$	1.520	1.230	.657	.343
$Star_i$	1.340	1.160	.748	.252
Mean VIF	1.320			

Note: The variables of response and review are one month lagged.

A few endogeneity issues need to be taken into consideration when constructing the model. First, the inherent service quality and managerial expertise may affect the level of customers' satisfaction. A higher star rated and larger scale hotel is more likely to implement managerial response strategies and have better customer ratings. Two hotel characteristics (i.e., $Star_i$ and $LnSize_i$, representing hotel star class and logarithmic number of rooms respectively) are included to control for time-invariant hotel specific factors. One point to notice is that the star ratings and the number of rooms of some hotels may have changed in the time range of the data. However, due to unviability of the historical data, the analysis uses the data entry at the point of data collection and assumes these two attributes are time-invariant.

Second, there might also be unobservable contemporaneous quality investments (e.g., major improvement in service quality and firm capabilities) or service failure crisis that may cause changes in customer ratings and business response behaviour. These shocks may be reflected in the changing customer review characteristics. For example, a hotel undergone a major renovation may present improved quality of facilities, potentially leading to a hike in customer satisfaction and review ratings in the following period. Thus, the review descriptors with moving values mentioned above are used to account for the time-varying hotel specific factors.

In addition, the data has an inherent hierarchical structure, that is, the hotel–month observations are nested within groups (i.e., hotel). To be specific, level 1 data is the reviews and responses of hotel i in month t, and they are clustered at each hotel i which is the level 2 identifier. As previously discussed, there might be unobservable hotel level heterogeneity and possible time effects varying across hotels that affect ratings. To account for this issue and analyse the clustered data, a multilevel model is adopted which includes both fixed and random effects and allows for variation between groups and estimation of group effects (Gelman & Hill, 2007). Specifically, in the random part varying intercept for each hotel and varying slope for each time period are allowed. The two-level model is as follows:

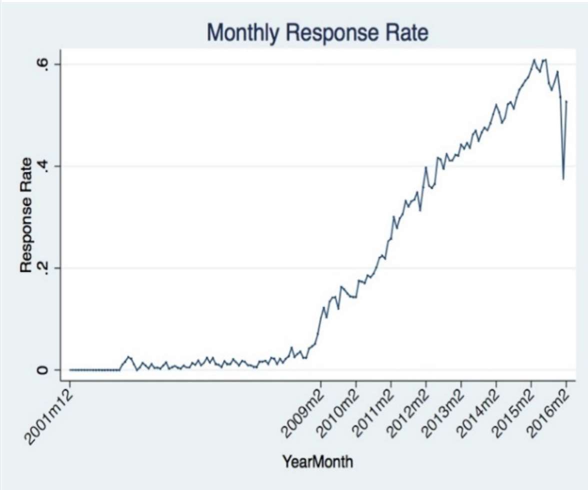
$$Rating_{i,t}=\beta_0+\varphi RESPONSE_{i,t\text{-}1}+\gamma RESPONSE_{i,t\text{-}1}\times REVIEW_{i,t\text{-}1}+\delta REVIEW_{i,t\text{-}1}+\tau_i+u_{0i}+u_{it}+e_{it}$$

where $RESPONSE_{i,t\text{-}1}$ is the vector of management response variables and $REVIEW_{i,t\text{-}1}$ is the vector of customer review variables of hotel i in year-month $t\text{-}1$; τ_i represent the time invariant hotel factors; u_{0i}, u_{it}, and e_{it} capture the random effects of hotel i and time t on ratings as well as observation-level residuals. φ and γ are the interested parameters which capture the effects of management responses on subsequent customer ratings and the moderating effects of review nature on management response influence.

4.4 Data Analysis and Results

4.4.1 Descriptive statistics

Along the timeline, both customers and firms become more engaged in the online review network (see Figure 4-2). The volume of online customer reviews and management responses both increased annually, and the growth started to accelerate around the year 2010. The monthly response rate in the sample shows a steady rise over past 15 years with an obvious boost after the year 2009. Among the 1,037 hotels with customer reviews displayed on the website, 735 hotels (70.88%) provide online managerial responses, with an overall response ratio of 43.61%.



Notes: The drop in 2016 is due to data collection. The data was collected in March 2016 and refined to the period before the end of February 2016.

Figure 4-2. Review and response volume over time

Figure 4-3 presents the percentage of customer ratings on a scale of 1 to 5 for each year over the sample period. There is an increase in the proportion of 5-score reviews while a decrease in that of 1- or 2-score reviews. Neutral reviews with a rating of 3 and 4 make up a relatively fixed percentage of the reviews over the years. This may reflect possible changes in customer behaviour and quality improvement in the hotel sector.

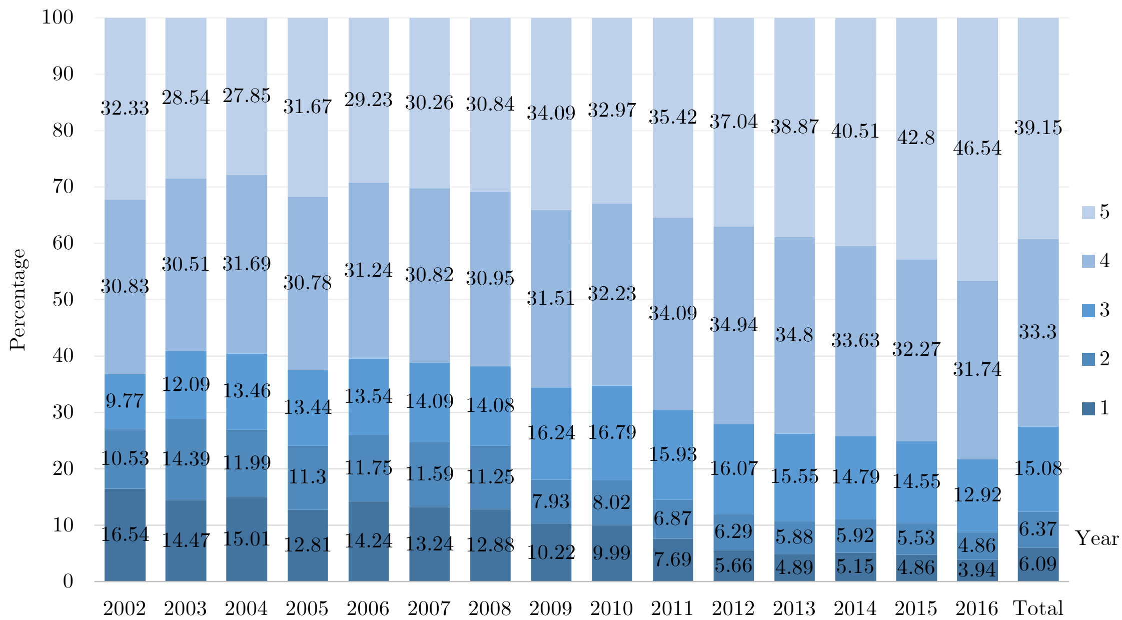


Figure 4-3. Distribution of ratings over time

Management response provision varies with hotel star class and overall customer ratings. As shown in Table 4-4, hotels of higher star class tend to be more active in responding to reviews compared to those lower rated. Over 90% of 4- and 5-star hotels offer responses to more than half of online reviews they receive. About 77% of 3-star hotels are responding but the response rate almost halves compared to that of luxury hotels. The numbers of responding hotels are smaller at the 3-star level or below and a limited number of responses are provided. In terms of customer ratings, hotels with an average score of 4 and 4.5 are most active in responding to online reviews (87.43%). Incentives for 3-score hotels are reduced (67.47%) and a lower response rate is observed (31.81%). Very few hotels rated below 3 by customers are involved in online responding, and the response ratio is much lower than that of higher rated hotels. A possible reason for above-average hotels being more engaged in responding may be that these hotels are willing to make the effort to get high ratings on the platform by doing what the community values. Besides, the management expertise of high-end businesses is more likely to be of a high standard, where online reviews are treated more seriously and providing managerial responses is a strategic policy.

Table 4-4. Descriptive statistics of response provision

Hotel		Number of hotels	Percent 1 (%)	Number of hotels with response	Percent 2 (%)	Volume of reviews	Percent 3 (%)	Volume of response	Percent 4 (%)
Star	5	110	10.61	102	92.73	114409	15.81	64310	56.21
Class	4/4.5	274	26.42	254	92.70	328760	45.43	179511	54.60
	3/3.5	363	35.01	279	76.86	221703	30.64	60930	27.48
	2/2.5	135	13.02	67	49.63	44125	6.10	9520	21.58
	1/1.5	13	1.25	5	38.46	2491	.34	168	6.74
	0	142	13.69	28	19.72	12192	1.68	1172	9.61
	Total	1037	100	735	70.88	723680	100	315611	43.61
Overall	5	44	4.24	28	63.64	24604	3.40	13316	54.12
Customer	4/4.5	509	49.08	445	87.43	487854	67.42	247248	50.68
Rating	3/3.5	292	28.16	197	67.47	164917	22.79	52452	31.81
	2/2.5	160	15.43	59	36.88	43431	6.00	2558	5.89
	1/1.5	32	3.09	6	18.75	2874	.40	37	1.29
	Total	1037	100	735	70.88	723680	100.00	315611	43.61

Note: [1] The value 0 of hotel star class indicates unavailability of information on the website, either because the hotel is not star rated or it is not provided by the site.
[2] Percentage 1 is calculated as the number of hotels at a certain rating level divided by the total number of hotels; Percentage 2 shows the ratio of responding hotels to all hotels at the same rating level; Percentage 3 is given by number of reviews on hotels at a certain rating level divided by the total volume of reviews; Percentage 4 shows the response rate of hotels at the same rating level.

Our sample covers 1,037 hotels across 171 months. The monthly average customer rating (*Rating*) is 3.550 (SD = 1.074). The moving average of customer ratings (*ReviewValence*) has a mean value of 3.501 (SD = .855). Both two variables have negative skewness (-.831 and -.675), indicating reviewers are slightly more positive in ratings on this website. Overall, the mean monthly response ratio across all panels is about 22% (*ResponseRatio*, M = .221, SD = .371), and the average responding speed (*ResponseDays*) is 40 days after the review posts (M = 39.95. SD = 214.3). As for the textual content of managerial responses, on average hotels write 91 words in the responses (*ResponseLength*, M = 90.96, SD = 56.34), and the overall content similarity is at a low level (*ResponseStandardisation*, M = .087, SD = .061). The average sentiment score is .658 (*ResponseSentiment*, SD = .095), showing a relatively positive sentiment when writing responses.

The correlations among variables are checked using Pearson correlation coefficients. The correlation matrix presented in Table 4-5 shows that most variables are correlated at .05 significance level, with the highest correlation coefficient of .738. A positive relationship is found between *Rating* and *ResponseRatio* (r = .266, $p < .001$), as well as between *Rating* and *ResponseSentiment* (r = .225, $p < .001$). There is a negative relationship between *Rating* and *LnResponseDays* (r = -.116, $p < .001$), *Rating* and *LnResponseLength* (r = -.019, p < .001), and *Rating* and *ResponseStandardisation* (r = -.093, $p < .001$).

Table 4-5. Variable correlation matrix

	1	2	3	4	5	6	7	8	9	10	11
1.Rating$_{i,t}$	1										
2.ResponseRatio$_{i,t-1}$	.266*	1									
3.LnResponseDays$_{i,t-1}$	-.116*	.087*	1								
4.LnResponseLength$_{i,t-1}$	-.019*	-.170*	-.135*	1							
5.ResponseStandardisation$_{i,t-1}$	-.093*	.011	.212*	-.085*	1						
6.ResponseSentiment$_{i,t-1}$	.225*	.266*	-.005	-.087*	-.036*	1					
7.ReviewValence$_{i,t-1}$	.738*	.302*	-.116*	-.017*	-.109*	.262*	1				
8.LnReviewVolume$_{i,t-1}$	.230*	.405*	-.305*	.056*	-.186*	.164*	.257*	1			
9.ReviewVariability$_{i,t-1}$	.084*	.298*	-.153*	.064*	-.117*	-.011	.088*	.632*	1		
10.Star$_i$	.397*	.274*	-.035*	.060*	-.155*	.232*	.514*	.218*	.167*	1	
11.LnSize$_i$	.216*	.239*	-.159*	.023*	-.240*	.161*	.257*	.346*	.426*	.431*	1

Note: Variables 2–9 are one month lagged. * $p < .05$

4.4.2 Effects of management responses

Table 4-6 presents the results of estimation using the multilevel modelling approach. The results in column (1) demonstrate that response ratio (*ResponseRatio*, β = .029, p = .021) has a significant positive influence on ratings in the subsequent period. A 10% increase in the response rate is associated with a .0029 increase in later ratings. Second, response days are significantly and negatively related to future ratings (*LnResponseDays*, β = -.019, p < .001). A quicker average response speed, say 10 days shorter in time intervals between review and response posts, is associated with .019% increase in the next period's ratings. Moreover, a positive significant relationship is found between future ratings and sentiment (*ResponseSentiment*, β = .156, p = .003). A .1 increase in sentiment scores is related to a .0156 increase in ratings in the next period. In addition, it is observed that response text length (*LnResponseLength*, β = -.010, p = .272) and content standardisation (*ResponseStandardisation*, β = -.051, p = .651) have no statistical significance in influencing future ratings.

Table 4-6. Estimations of response effects

Rating	Multilevel model		
	(1)	(2)	(3)
Fixed effects			
$ResponseRatio_{i,t-1}$	.029**	.072***	.081***
	(.012)	(.019)	(.023)
$LnResponseDays_{i,t-1}$	-.019***	-.022***	-.022***
	(.004)	(.005)	(.005)
$LnResponseLength_{i,t-1}$	-.010	-.017	-.018
	(.009)	(.013)	(.014)
$ResponseStandardisation_{i,t-1}$	-.051	-.024	-.034
	(.112)	(.093)	(.104)
$ResponseSentiment_{i,t-1}$	.156***	.186***	.166**
	(.053)	(.066)	(.072)
$ReviewValence_{i,t-1}$	.766***	.782***	.780***
	(.019)	(.022)	(.022)
$LnReviewVolume_{i,t-1}$	-.015***	-.012**	-.005
	(.005)	(.005)	(.007)
$ReviewVariability_{i,t-1}$	-.014	-.042	-.037
	(.045)	(.046)	(.045)
$ResponseRatio_{i,t-1}$×$ReviewValence_{i,t-1}$		-.109***	-.102***
		(.028)	(.028)
$LnResponseDays_{i,t-1}$×$ReviewValence_{i,t-1}$		.007	.007
		(.007)	(.007)
$LnResponseLength_{i,t-1}$×$ReviewValence_{i,t-1}$		.028	.030
		(.019)	(.018)
$ResponseStandardisation_{i,t-1}$×$ReviewValence_{i,t-1}$		.216**	.211**
		(.089)	(.099)

$ResponseSentiment_{i,t-1}$×$ReviewValence_{i,t-1}$		-.122	-.150
		(.092)	(.096)
$ResponseRatio_{i,t-1}$×$LnReviewVolume_{i,t-1}$			-.016
			(.010)
$LnResponseDays_{i,t-1}$×$LnReviewVolume_{i,t-1}$			-.001
			(.003)
$LnResponseLength_{i,t-1}$×$LnReviewVolume_{i,t-1}$			-.001
			(.010)
$ResponseStandardisation_{i,t-1}$×$LnReviewVolume_{i,t-1}$			.005
			(.060)
$ResponseSentiment_{i,t-1}$×$LnReviewVolume_{i,t-1}$			.060
			(.045)
$LnSize_i$	-.005	-.003	-.004
	(.009)	(.009)	(.009)
$Star_i$	.058***	.060***	.060***
	(.010)	(.010)	(.010)
Intercept	.866***	.800***	.781***
	(.099)	(.115)	(.129)
Random effect variances			
Hotel level	.014***	.014***	.014***
	(.004)	(.004)	(.004)
Monthly review/response level	.172***	.172***	.172***
	(.007)	(.007)	(.007)
Month Time effect	.000***	.000***	.000***
	(.000)	(.000)	(.000)
N	23277	23277	23277
LogLikelihood	-13071.546	-13036.124	-13032.878
Chi-square	3876.455	4163.606	4279.897

Note: The independent variables of response and review are one month lagged. All estimations have robust error terms clustered at the hotel level. Standard errors in parentheses. * $p < .1$, ** $p < .05$, *** $p < .01$

As for the review variables, results indicate that there is a positive effect of the review valence averaged based on ratings in all the previous periods on later ratings (*ReviewValence*, β = .766, $p < .001$). The cumulative volume of reviews in previous *t-1* periods has a negative impact on period-*t*'s ratings (*LnReviewVolume*, β = -.015, $p = .005$). Review variability has no statistically significant relationship with ratings (*ReviewValence*, β = -.014, $p = .757$). The control variable *LnSize* is not significant either ($p = .602$), yet star class of hotel seems to positively affect customer ratings (*Star*, $p < .001$). Furthermore, by interpreting the random effect parameters, it is found that only 7.5% of the variance in ratings can be attributed to differences between hotels, while the remaining is attributed to differences within individual hotels.

Next, interaction terms of centralised responses and review variables are included to examine whether the effects of management responses on ratings are moderated by review characteristics (see columns (2) and (3) in Table 4-6). The positive effect of response ratio is negatively moderated by mean ratings in previous periods ($p < .001$). It suggests that response ratio has a weaker positive influence given a higher mean rating, while the marginal effects on ratings are larger given a lower mean rating. Besides, there is no moderating effect of review valence on the influence of response speed and sentiment on future ratings, implying the imperative of response speed and sentiment regardless of the mean ratings in previous periods. Moreover, no significant moderating effects of review volume on response effects are observed.

4.4.3 Robustness check and additional analysis

A few additional tests are undertaken to check the robustness of results and obtain further insights. First, the focal model is tested using OLS regression with standard errors clustered at the hotel level. Hotel dummies and month dummies are included to control for firm and time effects. The estimations in Table 4-7 Column (1) and (2) are similar to the base results. Response ratio and response sentiment have a significantly positive impact and response days has a significantly negative impact on future ratings. These effects can be moderated by the cumulatively averaged ratings in previous periods. Second, the median value of all the response attributes is used, except for response ratio, in a certain period, instead of mean value. Results are derived from both multilevel regression and OLS regressions. Estimations of the main effects of response attributes are consistent with the base results and the cumulatively averaged rating can moderate the effect of response ratio on future ratings.

Table 4-7. Estimations using different time window

Rating	Monthly data (OLS)		Weekly data		Quarterly data	
	(1)	(2)	(3)	(4)	(5)	(6)
Fixed effects						
$ResponseRatio_{i,t-1}$	.036**	.073***	.007	.012	.054***	.125***
	(.015)	(.021)	(.011)	(.015)	(.015)	(.031)
$LnResponseDays_{i,t-1}$	-.018***	-.023***	-.015***	-.018***	-.017***	-.018***
	(.004)	(.005)	(.003)	(.004)	(.004)	(.006)
$LnResponseLength_{i,t-1}$	-.005	-.012	-.030***	-.033***	.005	.003
	(.011)	(.014)	(.008)	(.009)	(.010)	(.018)
$ResponseStandardisation_{i,t-1}$	.218	.172	-.037	-.045	-.122	-.091
	(.134)	(.173)	(.092)	(.092)	(.110)	(.104)
$ResponseSentiment_{i,t-1}$	.160***	.210***	.249***	.273***	.135**	.131
	(.059)	(.075)	(.037)	(.041)	(.064)	(.097)
$ReviewValence_{i,t-1}$	.342***	.361***	.768***	.760***	.795***	.814***
	(.059)	(.061)	(.016)	(.018)	(.021)	(.022)
$LnReviewVolume_{i,t-1}$	-.016	-.016	-.015***	-.016***	-.013**	-.006
	(.010)	(.010)	(.004)	(.005)	(.006)	(.008)
$ReviewVariability_{i,t-1}$	.097	.052	-.012	-.024	.057	.028
	(.077)	(.079)	(.035)	(.036)	(.055)	(.059)
$ResponseRatio_{i,t-1}$×$ReviewValence_{i,t-1}$		-.096***		-.027		-.113***
		(.030)		(.025)		(.034)
$LnResponseDays_{i,t-1}$×$ReviewValence_{i,t-1}$		.016**		.015**		-.005
		(.007)		(.006)		(.007)
$LnResponseLength_{i,t-1}$×$ReviewValence_{i,t-1}$		.024		.021		-.006
		(.021)		(.015)		(.019)
$ResponseStandardisation_{i,t-1}$×$ReviewValence_{i,t-1}$		.040		.077		.105
		(.231)		(.110)		(.111)
$ResponseSentiment_{i,t-1}$×$ReviewValence_{i,t-1}$		-.172*		-.277***		-.235**
		(.102)		(.065)		(.116)
$ResponseRatio_{i,t-1}$×$LnReviewVolume_{i,t-1}$				.005		-.015
				(.010)		(.013)
$LnResponseDays_{i,t-1}$×$LnReviewVolume_{i,t-1}$				.001		.003

				(.003)		(.003)
$\text{LnResponseLength}_{i,t-1}\times\text{LnReviewVolume}_{i,t-1}$				-.001		.002
				(.007)		(.011)
$\text{ResponseStandardisation}_{i,t-1}\times\text{LnReviewVolume}_{i,t-1}$				-.014		.035
				(.051)		(.065)
$\text{ResponseSentiment}_{i,t-1}\times\text{LnReviewVolume}_{i,t-1}$				.036		.097*
				(.034)		(.052)
LnSize_i	1.388***	1.362***	-.004	-.005	-.011	-.010
	(.225)	(.375)	(.009)	(.010)	(.008)	(.008)
Star_i	.639***	.520	.056***	.057***	.053***	.057***
	(.106)	(.344)	(.010)	(.010)	(.009)	(.010)
Intercept	-4.721***	-4.268***	.881***	.929***	.668***	.556***
	(.738)	(.732)	(.085)	(.100)	(.120)	(.161)
Random effect variances						
Hotel level			.010***	.010***	.015***	.015***
			(.002)	(.002)	(.004)	(.004)
Weekly/Quarterly review/response level			.350***	.350***	.101***	.100***
			(.010)	(.010)	(.006)	(.006)
Week/Quarter Time effect			.000***	.000***	.000***	.000***
			(.000)	(.000)	(.000)	(.000)
N	23277	23277	67429	67429	9402	9402
LogLikelihood			-60847.095	-60812.608	-2927.393	-2892.436
R-square	.626	.627				
Chi-square			4326.164	4654.582	4537.767	5171.689

Note: Hotel and month dummies are included in the OLS regression (column 1 and 2). The independent variables of response and review are one week (column 3 and 4) or one quarter (column 5 and 6) lagged in the multilevel model. All estimations have robust error terms clustered at the hotel level. Standard errors in parentheses. * $p < .1$, ** $p < .05$, *** $p < .01$

In addition, in the main test, values of variables are aggregated at the monthly level. The one month lag is appropriate given that the average response timeframe across the panels is 40 days. A one-month time window is allowed for responses to be observed by prospective customers and reviewers. However, the median value of *ResponseDays* in the sample is 6.7 days, suggesting that typically a response can be displayed and thus observable within a week. Therefore, a relatively tighter time window is used to examine managerial response's impact on weekly ratings. Table 4-7 presents the estimation results with weekly data (Column (3) and (4)). The negative impact of response speed and the positive impact of response sentiment on future ratings remain consistent with the base results. However, response ratio no longer has a significant effect, while response length becomes significant in negatively influencing the next period's ratings. Moreover, the cumulatively averaged rating in previous periods can moderate the effects of response days and response sentiment on ratings. When the mean rating established in previous *t-1* periods is higher, the marginal negative effects of response days and marginal positive effects of response sentiment tend to be reduced. No moderating effects are discovered for the other review attribute. The results are further checked using a quarterly time window (Column (5) and (6) in Table 4-7) and the results are similar to the based results and the effect of response ratio is stronger while the effect of response length is more insignificant than the monthly estimations. These findings suggest that the significance and magnitude of response effects vary with time. Influence of response ratio may appear in a longer time window while response length may only be important in a short period.

4.5 Discussions and Implications

Our empirical examination reveals that higher response ratio, prompt response and showing positive sentiment in the response text would be favourable for enhancing future customer ratings. First, in line with findings in Xie et al. (2016), responding hotels are more likely to receive higher ratings in the future. Higher response ratio indicates a firm's continuous engagement in online responding, which is likely to enhance public recognition of a firm's effort to interact with customers. Second, the importance of response promptness on future ratings is recognised. Unlike Min et al. (2015) who report no influence of response speed on ratings, this study finds a strong association between the two variables. An increase in rating score in the next period tends to occur if the time intervals between the review post and response are shorter. Timely actions indicate that firms are checking and responding on a regular basis. It might be particularly important for the online complaint scenario as it suggests managerial efficiency in offering viable solutions. In addition, these two attributes potentially increase the probability of response visibility on the online review platform, leading to an influence on future customers' rating behaviour.

As for the response sentiment, improved ratings are achievable by showing more positive sentiment in the response texts. This finding is similar to the observations in Li and Song (2010) and Min et al. (2015) that accommodative and empathetic responses are effective in improving future customers' evaluation of firms. Firms that respond with a courteous tone of voice show their politeness and attentiveness to customers. Being supportive and sincere in responding to customer reviews is also helpful in building a healthier, more positive public image and goodwill. It is a demonstration of managerial professionalism in handling customer online feedback, leading to potential improvement in future ratings.

The magnitude of the above effects of responses is more pronounced when the overall customer rating is lower, but no significant moderating effects of review volume and variability on response influence are found. The positive effects of response ratio and sentiment are diminishing if the overall customer review sentiment is positive, but are more significant in an unfavourable rating situation. It implies that the scope of response provision should not be limited to certain types of review. More attention could be paid to unfavourable reviews. Firms can take advantage of the online communication channel to perform service recovery by responding in a positive, polite and responsible way. It may help reverse the destructive effects of negative wording and give prospective customers a good impression.

In addition to the above findings that conform to expectation, there is no sufficient statistical evidence to support the hypotheses that response length and standardisation have a significant impact on future ratings. Although longer and tailored responses are potentially more informative and demonstrate managerial willingness to write proper replies, the presumed uncertainty reduction advantage associated with these two attributes is not discovered. One possible reason is that the review site shows only the first three lines of a response if customers do not click on the 'more' link to unfold the hidden messages. In this case, potential customers and reviewers seem to be less affected by the detailed textual content of managerial responses. They may pay little attention and hardly react to how long a response message is and whether the content of responses is similar. The specific textual content and details may be more accentuated in influencing existing customers' future satisfaction but not that of bystanders.

Our findings affirm the position of the social information processing theory, which explains interpersonal relationship development, impression formation and social cognitive processes in a computer-mediated communication (Walther, 1992). In a text-only communicating environment where non-verbal cues (e.g., gestures, facial expressions, and body position) are inefficient, online communicators would adapt to any socially revelatory information to draw inferences and form an impression (Walther, 1992; Walther, 1996). The response content, response time, style in writing and emotional context are effective information to advance firm–customer relationships over time (Walther, 2008). In addition, given that customers are exposed to multiple sources of influence (Walther, Liang, Ganster, Wohn, & Emington, 2012), it is found that the non-textual cues embedded in response behaviour (i.e., response ratio, speed, sentiment) are more powerful in influencing future ratings than text-related dimensions (i.e., response length and content standardisation).

4.5.1 Research contributions

This chapter offers several contributions to the marketing and strategy research. First, the imperative of firm engagement in online communications is recognised by demonstrating how response strategies have an impact on future customer ratings. Instead of taking the regime change from responding to responding as a treatment, this study focuses on the continuous and dynamic impact after the initial response. The empirical test documents some evidence for the continuous impact of responses in the long run, which offers new insight into the ongoing effects of response arrangement on the subsequent rating behaviour.

Besides, this work goes beyond the traditional emphasis on how the existence or the volume of managerial responses makes a difference to future customer rating by focusing on the detailed nature of managerial responses such as timeliness, length, standardisation and sentiment of response, which have been overlooked in the current literature. This study examines these response attributes simultaneously and provide new insights into the quality aspects of responses and their influences on subsequent customer ratings. The investigation into specific characteristics of responses enables us to address the complexity of managerial responses and identify the mechanism for engendering impact on future ratings.

Moreover, considering the interplay between online reviews and managerial response, their joint effects on future ratings are examined. The study attempts to figure out how to respond given the different nature of reviews such as valence, volume and variance (see Babić Rosario et al., 2016). Deviated from a practical bias towards using managerial responses to deal with consumer complaints, this study depicts a broader picture of managing online reviews taking all types of reviews into account. This has strategic significance as it suggests that the effectiveness of response strategies varies in different eWOM scenarios.

In addition, this chapter takes advantage of a large number of online reviews and responses available online to study how managers and customers actually interact on the Internet-enabled platform. Although earlier research relying on an experimental design (e.g., Min et al., 2015; Sparks et al., 2016) has its merits in controlling over extraneous variables, such scenario-based study creates an artificial situation and is subjected to human error, which raises concerns for generalisability and validity and of the findings (Xie et al., 2016). In this sense, the data-driven research using real hotel reviews and manager responses delivers field evidence on the efficacy of management responses. Furthermore, given the unstructured nature of response data, the textual content of management response is analysed using natural language processing. It offers a new angle to look into the multifaceted managerial response issue, which yields new insights into handling different types of reviews.

4.5.2 Practical implications

Our work offers several implications for practice. First of all, managerial response opens opportunities for business to engage in customer relationship management. The responding behaviour of firms revealed on the online communication platform can influence the impressions and perceptions of potential customers and reviewers. Accordingly, providing online managerial responses should be a part of the strategic process to maintain ongoing relationships with customers (Park & Allen, 2013). This reemphasises the need for firms to allocate resources (Maritan & Lee, 2017) to take advantage of managerial responses to make a continuous impact on customer rating behaviour.

To improve future ratings, firms need to respond with a higher ratio, promptness and positive sentiment. Decisions on various responding strategies should depend on the nature of customer reviews received by the firm. When previous customer opinions lean towards negative, firms need to raise the response ratio and demonstrate a more positive sentiment in the response. It helps to mediate the spread of unfavourable effects of bad eWOM by creating the image of a responsible firm. In contrast, if customer reviews accumulated from previous periods are more positive, firms could potentially respond less and tone down the language as the marginal effects on future ratings are less significant.

There are, however, a few points worth noting. First, the accuracy of information in the managerial responses is critical to the validity of signals (Ippolito, 1990). An overstatement of the service quality would lead to firms struggling to meet the high expectation of customers formed by the positive signals. More importantly, customers who recognise the information signals that businesses sent to the public are strategically chosen may respond to the inaccurate information against their own consumption experience by rating lower. Second, it is critical to determine how fast a firm should respond to complaints. Immediate actions may not always be ideal. Firms may fail to completely address issues raised in customer reviews, which may generate the impression that firms only make a perfunctory effort. The right timing will depend on the nature of concerns expressed in the reviews and possible solutions to solve the problems. Furthermore, there is a possibility that firms overuse upbeat language in response messages, which may have an adverse effect. The overwhelmingly positive response may be perceived as insincere marketing behaviour. Customers may question whether the responses with great unction have substantial implications for a firm's offline investment. It suggests that managers should speak positively but not rhetorically.

4.5.3 Limitations and directions for future research

Several limitations of this study must be borne in mind. First, this chapter focuses on firms' online activities but are not able to specifically capture all effects of offline unobservable variables such as service quality, renovation and time-varying managerial expertise. Future research could add in new variables to further check the robustness of estimations in current settings. Second, this study uses reviews and responses of London hotels and organises data at the monthly level in the main test, so the results should be interpreted with caution. Future studies could break down the geographical, platform and timeframe constraints to gain additional insights. For example, the model can be validated using data from different markets and online review platforms. Moreover, this study does not segment the analysis based on organisational characteristics (e.g., independent or chain hotels, different star classes). Future work could gain additional insights into the magnitude of response effects by targeting specific types of business. In addition, it may also add value to the strategic action frame proposed in this study by defining the boundaries of each response attribute to achieve optimal outcomes.

CHAPTER 5

THE STRATEGIC VALUE OF BUSINESS ONLINE PRESENCE AND ACTIVENESS: STIMULATING CUSTOMER ENGAGEMENT IN ONLINE COMMUNICATIONS

5.1 Introduction

Supported by Web 2.0 (O'Reilly, 2007), information creation and exchange in social media are increasingly common in the digital world (Kaplan & Haenlein, 2010). In this participatory environment, many more customers are willing to participate in the online communications to share their experiences with firms and other members of the community. Such interactive activities take place in various forms, including blogging, word-of-mouth communications, writing reviews, and recommendations (Van Doorn et al., 2010). The massive content customers write can be a useful information source for firms (Kozinets, 2002) in developing and improving businesses' dynamic marketing capabilities (Barrales-Molina et al., 2014). More importantly, engaged customers become online word-of-mouth advertisers for businesses, and this imposes a profound impact on the deeper level of firm–customer relationships and long-term business performance.

The power of sharing is determined by the breadth and depth of customer engagement, which has become an essential feature of businesses. For example, the volume of user-generated content (UGC) for a product/service or the number of users in an online brand community serves as an indicator of a brand's or a product/service's popularity (Proserpio & Zervas, 2017). Such popularity may attract wider attention, potentially leading to increased recognition and sales (Tirunillai & Tellis, 2012), according to the social influence network theory (Friedkin, 1998). This theory posits that the social influence created by online traffic and the propagated information in a social network is pervasive in shaping individuals' attitudes, cognition and behaviour (Iyengar, Van den Bultem, & Valente, 2011; Kurt, Inman, & Argo, 2011). Management researchers regard such social influence among fellow consumers as one of the primary factors affecting consumers' choices of products (e.g., Kurt et al., 2011; Wang, Aribarg, & Atchadé, 2013), purchase intentions (e.g., Fang et al., 2013; Zhang et al., 2014), perception (e.g., Cheng & Ho, 2015; Lee et al., 2015), and online communication behaviour (e.g., Goes et al., 2014; Sridhar & Srinivasan, 2012; Zhang et al., 2011). A wider scope of customer engagement can amplify the crowd's voice on the Internet and hence strengthen the social influence, with a greater number of customers acting and participating in the network (King et al., 2014).

The substantial influence of customer engagement on business performance calls for marketers to expand network scale to acquire market knowledge, competing for a portion of public attention, improving online reputation, retaining and satisfying customers and creating synergistic effects (Chang, Yu, & Lu, 2015). To encourage customers to voice their opinions, companies are now acting in social media. For instance, many firms initiate and manage fan pages on social networking sites to breed online brand community. The virtual online community in the computer-mediated social gathering context fosters customers' engagement in the network (Shriver et al., 2013), leading to increased trust and network effects of the social influences on connected people reciprocally (Fang et al., 2013; Shoham, Moldovan, & Steinhart, 2017). In addition to interactions among community members, we also observe purposeful marketing posts by firms and a growing number of online firm–customer conversations such as chatting with customers and responding to customers' online posts on review platforms or discussion forums. Marketing researchers and practitioners have recognised that social media are becoming a desired and efficient channel connecting consumer and marketers (Schniederjans et al., 2013; Hollebeek, Glynn, & Brodie, 2014). It is useful for disseminating information and engaging customers, through which companies impose influence on customers' behaviour (Evans, 2010).

Given the business impact of customer engagement and thus the importance of business strategy to encourage customer engagement behaviour via social media, the key questions that motivate this study are, "To what extent can business social media presence affect customer engagement behaviour?" and "How active businesses should be online to positively influence customer engagement behaviour?" A review of the literature does not give a clear answer. The first question is concerned with the antecedents of customer engagement, for which prior studies primarily focus on the customer, firm or context specific factors (Van Doorn et al., 2010). King et al. (2014) illustrate that in the current body of knowledge of the antecedents of electronic word-of-mouth (eWOM) participation, there is a need to know how firms can foster reviews and reviewers. In particular, little is known about whether firm engagement in social media activities, in other words, the firm–customer interaction, is also a motivational driver for customer engagement in eWOM. The latter question is of greater interest and higher practical relevance. Research on firms' strategic use of online social sites is in an early stage (Goh et al., 2013), with current efforts devoted to debating whether to engage in social media activities and evaluating their economic value. However, little attention has been paid to the efficacy of firms' social media efforts in affecting customer engagement rather than purchase behaviour (Lamberton & Stephen, 2016). Discussion is also lacking on business online activeness after strategic regime change (i.e., adoption of social media strategy) and how such activeness continuously impacts customers' engagement behaviour.

This chapter examines the behavioural value of business presence and activeness in social media to stimulate customers' engagement in the online network. Specifically, this study uses the action (i.e., the commencement of managerial responses on the review platform) and visibility (i.e., the degree to which customers observe the managerial response messages) of online managerial responses to customer reviews as a proxy for firms' social media presence and activeness, respectively. Customer engagement behaviour is represented by community members' participation in the online review posting. Using review and response data on 1,024 London hotels over a 15-year period, this research examines to what extent the action of online managerial response and the visibility of those responses can affect the number of customer reviews posted on the online review platform.

To be more specific, first, this chapter investigates whether, in addition to individual specific determinants (see Hennig-Thurau et al., 2004), customers' engagement in writing online reviews is also influenced by firms' presence in the online community network. In doing so, the analysis first tests if the volume of customer reviews is different between firms with managerial responses displayed online and others without. By using a propensity score matching (PSM) technique, a treated (responding firms) and a control (non-responding firms) group are created within an online customer review platform to estimate an average treatment effect of adopting the managerial response policy. The results demonstrate that responding firms have a greater number of customer reviews compared to non-responding firms. This finding reveals a firm-leading influence on customer engagement behaviour, which contributes to the engagement literature by studying firm engagement and identifying it as an additional motivational driver for customer engagement. This also contributes to the emerging customer engagement marketing theory (Harmeling, Moffett, Arnold, & Carlson, 2017) by offering empirical evidence for the effectiveness of such marketing efforts.

Second, special attention has been paid to the after-treatment ongoing effect since the regime change to identify the implicit intervening process in influencing consumers' mind-sets (Srinivasan, Vanhuele, & Pauwels, 2010). It is discovered that the visibility of managerial efforts in social media affects customers' participation in the longitudinal dimension. In the test of managerial responses and review volume, two determinants of response visibility are considered, the ratio of response to review and response speed, which can potentially affect the response visibility after responding firms start to reply. The results of the multilevel model estimations demonstrate a positive influence of response speed on subsequent review volume, but no significant effect of response ratio. This provides practical insights for managers into how to increase response visibility and thus potential for influencing customer engagement. This finding contributes to the social media marketing literature by emphasising business activeness in social media and the continuity and consistency of relevant practices to encourage more customers to engage in online communications.

5.2 Related Literature and Hypothesis Development

5.2.1 Social media marketing and customer engagement marketing

In response to the social sense of business, marketers are gaining enthusiasm for capitalising on the social context and social influence for marketing activities (Yadav, De Valck, Hennig-Thurau, Hoffman, & Spann, 2013). Social media marketing is defined by Felix, Rauschnabel and Hinsch (2017, p. 123) as "an interdisciplinary and cross-functional concept that uses social media (often in combination with other communications channels) to achieve organisational goals by creating value for stakeholders". Kozinets, De Valck, Wojnicki, and Wilner (2010, p. 71) describe social media marketing as the "intentional influencing of consumer-to-consumer communications by professional marketing techniques". Social media marketing takes many forms, such as initiating fan pages on social networking sites and responding to customers' reviews on review platforms.

Current literature on social media marketing efforts and effectiveness mainly focuses on economic outcomes. Previous studies have documented that marketers play a persuasive role in social media, and marketer-generated content can affect customers' purchase behaviour (Goh et al., 2013). A positive association between social media marketing and purchase intention/expenditure (e.g., Kim & Ko, 2012; Kumar et al., 2016; Gong, Zhang, Zhao, & Jiang, 2017) results from the increased marketing capabilities built upon social media resources (Wang & Kim, 2017). Apart from driving revenue generation, such networking strategy is also powerful in brand management (Gensler, Völckner, Liu-Thompkins, & Wiertz, 2013). Godey et al. (2016) find that social media marketing favourably influences brand equity, especially brand awareness and brand image, as well as customers' behaviour towards the brand such as loyalty and preference. The creation and spread of firm-to-consumer social messages effectively enhance brand awareness, consideration and preference and attract new customers (De Vries, Gensler, & Leeflang, 2017).

In addition to exchange-related aspects, behavioural consequences of social media marketing efforts also are evident, particularly with regard to customers' engagement in online communications. Customer engagement is "a behavioural manifestation toward the brand or firm that goes beyond transactions" (Verhoef, Reinartz, & Krafft, 2010, p. 247), and "a multi-dimensional concept comprising relevant cognitive, emotional, and behavioural dimensions", varying in different contexts (Hollebeek et al., 2014, p. 152). Voluntary participation in social media can be both passive (i.e., reading the content generated by others) and active (i.e., creating content and sharing opinions) (Ashley & Tuten, 2015), which benefits the business, the brand and/or customers (Dong & Sivakumar, 2017). Incentives for customer engagement behaviour are multifaceted, involving customer-, firm- and context-related factors (Van Doorn et al., 2010). The literature documents that firms may be motivational drivers for customer engagement, mainly stemming from brand characteristics, venue/channel support, information environment and incentive rewards (see Van Doorn et al., 2010). However, very few studies have paid attention to firm–customer conversations and firm-generated content, and the findings of these studies are mixed (e.g., Kumar, Bhaskaran, Mirchandani, & Shah, 2013; De Vries et al., 2017). In fact, the social media marketing effort may also play a role in "chang[ing] customer engagement states—including their levels, intensities, and complexity" (Bolton, 2011, p. 273). Harmeling et al. (2017) define customer engagement marketing as "a firm's deliberate effort to motivate, empower, and measure a customer's voluntary contribution to the firm's marketing functions beyond the core, economic transaction" (p. 317). The objective of this marketing strategy is to motivate customers to actively participate and contribute to the marketing activities as "pseudo-marketers" (Harmeling et al., 2017, p. 312). Given the aim is to motivate customer engagement behaviour, it raises questions of how to motivate and how effective the strategy is. Nevertheless, little attention has been paid to firms' engagement in generating content in social media, and there is a lack of empirical evidence showing the effectiveness of firm–customer interactions on consumers' participation behaviour in online communications.

With respect to the nascent area of managerial responses to customer reviews, a handful of studies seem to suggest the potential impact of providing managerial responses on review volume. Ye et al. (2010) explore the impact of managerial responses on the volume of subsequent customer reviews. It applies a difference-in-difference approach to the customer review and management response data by matching hotels across two online review platforms. Comparing the volume of reviews before and after the first response, they find a positive impact of providing responses on review volume, but such influence diminishes if no further responses are provided. Using a similar cross-platform setting, Proserpio and Zervas (2017) touch on the impact of management responses on review volume when they discuss the mechanism for response affecting review ratings. They find an increase in review volume—especially the number of positive reviews—after hotels start to respond. Chevalier et al. (2017) present similar findings, demonstrating that managerial responses can stimulate customers' reviewing activities, particularly critical reviews. Furthermore, by testing a panel model, Xie et al. (2016) find that managerial responses can lead to an increase in the volume of subsequent consumer reviews. They attribute the increased number of consumer word-of-mouth to the online firm–customer interactions.

5.2.2 Behavioural effect of business social media presence and activeness

The review of previous studies reveals two gaps in the literature. One is the inconclusive discussion on the effectiveness of firms' social media presence in influencing customer engagement behaviour by firms engaging in online communications. The second gap is a lack of assessment of the business activeness of social media activities in making such behavioural effects. Nevertheless, examining the efficacy of firms' social media efforts is important because of the public nature of managerial responses and thus the potential influence on other consumers and potential reviewers.[5] Therefore, this study examines the effect of social media presence and activeness on customer engagement behaviour. In particular, it proposes that providing managerial responses may stimulate more customers to write reviews online, by discussing why reviewing activities may change with the introduction of managerial responses and how the visibility of managerial responses plays a role in a longer term.

The analysis in this study is closely related to but different from the aforementioned studies with regard to online managerial responses in several ways. First, a within-platform matching strategy is adopted instead of a cross-platform identification strategy. Drawing on data from a single review platform by matching responding and non-responding hotels based on their characteristics, it mitigates concerns about the comparability of different platforms in terms of their popularity, policy and the like, as well as the endogeneity issues caused by underlying hotel quality. Second, prior studies tend to estimate effects before and after the policy change rather than the long-term effect. This study focuses on how the visibility of responses on the platform makes an influence after firms start to respond. Two responding styles, i.e., response ratio and speed, which may affect response visibility, are investigated with regard to their continuous impact on review volume in a longer term.

5.2.2.1 Provision of managerial responses

The provision of managerial response in online review platforms may potentially affect customer engagement behaviour in several ways. First, the business-to-customer conversations establish firms' social media presence in the online virtual community, which is an indicator of business adoption of social media strategy. This is a behavioural manifestation of businesses listening to customers' opinions and acknowledging the associated efforts in writing the reviews. Such tactics explicitly signal to customers that firms are willing to listen (Proserpio & Zervas, 2017), which enhances consumers' inferences about the business' trustworthiness (Sparks et al., 2016). As a result, observing the firm–customer online interactions may inspire customers to voice with an expectation of that their opinions will be heard and responded to by the service

[5] In this study, the term 'potential reviewer' refers to existing customers who have not yet written reviews of their most recent service experience, regardless of whether they have written reviews before. Customers who have written multiple reviews for a hotel can be identified as returning customers. They may have strong preferences for the hotel and are more likely to write reviews, but this does not mean they will write reviews for the hotel again.

provider (Gu & Ye, 2014); in other words, engagement is potentially strengthened (Higgins & Scholer, 2009).

In addition, the content of conversations may provide incentives for expanded customer engagement. The online review-response establishes a communication channel connecting firms and customers and diffusing information (Felix et al., 2017). Management teams often acknowledge customers' word-of-mouth contribution, praise or distress, and promise to address the raised issues or sometimes offer offline benefits or compensation (Davidow, 2003). On the one hand, potential reviewers may be motivated to take advantage of the online and cost-effective medium with an expectation of their efforts being acknowledged, problems being solved or additional benefits being offered (Gu & Ye, 2014). On the other hand, the functional and social benefits may also prompt consumers to engage in order to obtain information, establish an interpersonal relationship with firms, and fulfil a social need (Homburg et al., 2015; Buechel & Berger, 2018). Furthermore, marketer-generated content can be viewed as advertisements, and the content of such exogenous word-of-mouth (Godes & Mayzlin, 2009) with a deliberate attempt to market the product or service may raise public scrutiny. Customers may react to the authentic or exaggerated information supplied by firms in the online conversations to a greater extent, leading to a higher propensity to engage and speak out. Therefore, it is hypothesised that:

Hypothesis 1: Firms providing online managerial responses have higher volume of customer reviews than non-responding firms.

5.2.2.2 Visibility of managerial responses

In addition to the adoption of social media marketing, the policy execution plays a more important role in a longer term. In the online social environment, the continuity and activeness of a business' online presence are manifested by the visibility of firm–customer interactions. In this study, visibility is defined as the extent to which potential reviewers are likely to observe online managerial responses. Higher visibility implies managerial responses are more easily observed by review readers, while lower visibility indicates that managerial responses are less observable and hence less likely to affect customers' review behaviour (Proserpio & Zervas, 2017). The visibility issue has high practical relevance. Customer reviews are displayed on the site in reverse chronological order, with the most recent appearing first and each page only displaying five reviews.[6] Many reviews are generated on the website every day, pushing older reviews and the associated responses to later pages and making them less observable than those on the first few pages (De Vries, Gensler, & Leeflang, 2012), given the fact that customers hardly ever go beyond the first few pages (Pavlou & Dimoka, 2006).

The visibility of managerial response is important in influencing customer engagement behaviour. The online visibility of the management team performs a symbolic function (Enz & Grover, 1992), signalling that the firm is active in embracing and managing customers' comments. It indicates the firm is devoting efforts to maintaining an interactive relationship with their customers, which is an essential element of management effectiveness. Moreover, the position of managerial responses is critical to determine its noticeability and hence the attention it is able to attract from review readers (De Vries et al., 2012). Two influential factors of response visibility are identified. One is the response ratio, measured as the number of responses divided by the number of customer reviews. A higher response ratio indicates that more managerial responses, in relationship to the number of reviews, are displayed on the site, resulting in a greater likelihood that customers will see them (De Vries et al., 2017). The second factor is the speed of response, measured by the average days between the date of the review posts and that of the associated responses. Timing and speed of response have a substantial influence in managing complaints and improving trust (e.g., Davidow, 2003; Homburg & Fürst, 2007; Sparks et al., 2016). To compete with the rapid update of reviews, quicker responses increase the probability of responses being displayed and visible on the first few pages (Wang & Chaudhry, 2018) and thus influencing potential reviewers. Hence, it is reasonable to hypothesise:

Hypothesis 2a: The ratio of online managerial response to customer review is positively associated with the future volume of customer reviews.

Hypothesis 2b: The speed of online managerial response is positively associated with the future volume of customer reviews.

[6] The review site has changed the number of reviews displayed on each page. At the time of data collection (March 2016), each page had ten reviews. Now there are five reviews per page.

Based on the research questions and hypotheses, Figure 5-1 outlines the research framework for the current study. The effects of managerial responses on review volume are tested from two aspects. The first is the relationship between the provision of managerial response and the review volume (H1). A second aspect is concerned with response visibility, which is described by the response ratio and the response speed. H2a and H2b are developed to test for the association between these two visibility descriptors and future review volume.

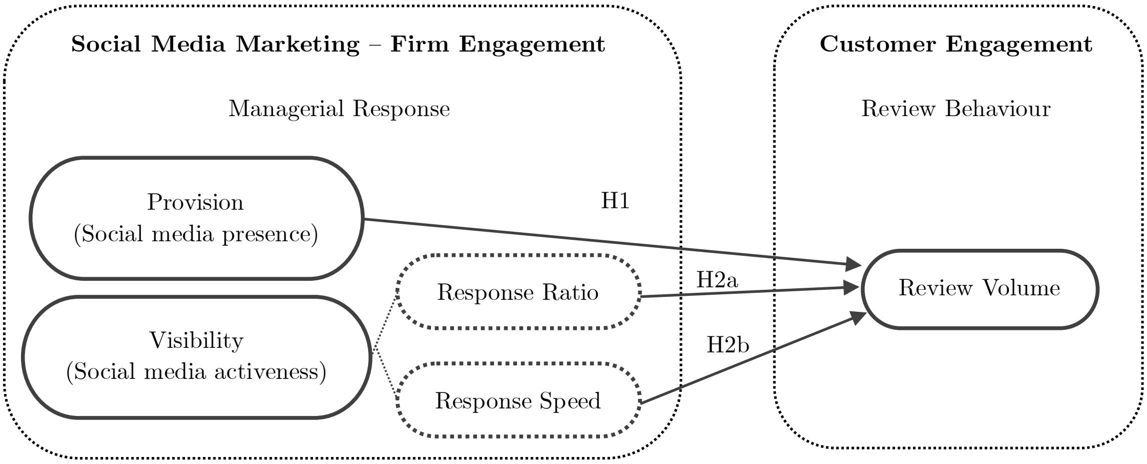


Figure 5-1. Research framework

5.3 Data

This study uses the data set collected from 1,063 hotels in London (as described in Section 1.5.3 of Chapter 1). The data is cleaned for subsequent analysis. First, hotels with zero reviews are removed (i.e., 26 hotels) as no observations of review and response are available to study the firm and customer engagement behaviour. Hotels that are closed at a later stage of the sample period are also excluded, leading to a loss of 13 more hotels. This is because it is hard to determine the causes of changes in online posting behaviour of both customers and service providers, which may bias the results. The final sample contains 1,024 hotels over about a 15-year period from January 2001 to February 2016. Although not all hotels in the sample appear on the site at the same time, the cut-off date for the review posts is the end of February 2016.

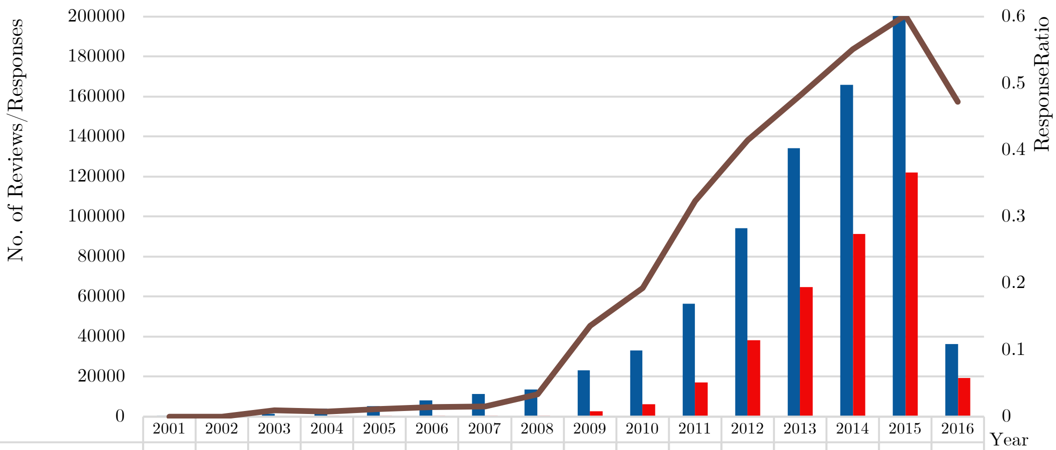
Table 5-1 presents the summary statistics of the data. Among the 1,024 hotels, 739 hotels (72.17%) have provided at least one response in the sample period. High-end hotels are most active in responding to online reviews (about 93% for both four-star and five-star classes). High customer-rated hotels (mean rating greater than or equal to 3) also actively engage in online managerial response, particularly those rated 4 or 4.5. This indicates better managerial expertise and the implementation of social media strategies in these firms. This observation is important to the subsequent analysis in terms of determining the factors endogenous to responding. Furthermore, in the sample, service providers respond to 45.93% of online customer reviews on the site. Along the timeline, there is a clear upward trend in writing online reviews and responses (see Figure 5-2). Many hotels in the sample received their first reviews on this platform in 2003, while most started to respond around 2009. As evident in Figure 5-2, the overall response ratio surges in the year 2009.

Table 5-1. Summary statistics

Star class	Number of hotels	Number of responding hotels	Mean rating	Number of hotels	Number of responding hotels
5	110	103 (93.64%)	5	44	29 (65.91%)
4/4.5	274	254 (92.70%)	4/4.5	506	445 (87.94%)
3/3.5	361	282 (78.12%)	3/3.5	289	199 (68.86%)
2/2.5	133	67 (50.38%)	2/2.5	154	60 (38.96%)
0/1/1.5	146	33 (22.60%)	1/1.5	31	6 (19.35%)
Total	1024	739 (72.17%)	Total	1024	739 (72.17%)

Note: The null value of star class is due to unavailability of this information on the website. Mean rating is the overall customer rated score, which is round to the nearest .5.

The main analyses are split into two parts. First, drawing on the whole sample, it explores whether hotels providing responses receive more customer reviews. To be specific, the test estimates the average treatment effect of responding on daily review volume by matching responding and non-responding hotels. Next, the analysis focuses on the responding hotels to examine whether the response visibility affects review volume over time using a panel approach.



	2001	2002	2003	2004	2005	2006	2007	2008	2009	2010	2011	2012	2013	2014	2015	2016
Number of Reviews	1	132	1,261	3,971	5,238	8,050	11,172	13,509	23,208	32,939	56,310	94,233	134,113	165,813	200,952	36,349
Number of Responses	0	0	0	8	11	65	108	299	2,696	5,969	16,992	38,269	64,594	91,223	122,097	19,286
Response Ratio	0	0	0.95%	0.73%	1.15%	1.40%	1.50%	3.28%	13.60%	19.30%	32.32%	41.49%	48.20%	55.05%	60.16%	47.22%

Note: The year of review and the year of its associated response may be different. Yearly distribution of response ratio is based on the year when the reviews were posted. The drop of review/response number in 2016 is due to data availability. The data was collected in March 2016 and refined to the period before the end of February 2016.

Figure 5-2. Number of reviews and responses

5.4 Effects of Responding: Matching Responding and Non-Responding Hotels

The first part of the analysis explores the association between providing online managerial responses and the number of customer reviews posted on the platform. The dependent variable is *DailyReviewVolume*, measured as the daily number of reviews each hotel receives during the period of its presence on the website. This duration (measured in days, denoted as *Age*) covers the period from the date of a hotel's first review to the cut-off date (the end of the sample period). The reason for not using the total number of reviews is that hotels appear on the website at different times, either because of market entrance or business listing on the site, which may affect the aggregated number of reviews. On average, hotels in the sample receive .344 reviews each day, with a minimum value of .0002 and a maximum value of 4.886 (the 25th, 50th, and 75th percentiles are .054, .140, and .409).

A dummy variable *Response* is created to indicate the provision of managerial responses of a hotel. This is the key explanatory variable, which takes the value of 1 if a hotel has provided at least one online response in the sample period. In addition, four variables related to hotels are used in the matching process and included in the daily review volume regression. These hotel specific factors may concern both the number of customer reviews and the probability of providing managerial responses. First, service quality may relate to the attractiveness of hotels and thus the number of online reviews. According to the insight from the summary statistics, highly rated hotels are more likely to respond to online reviews. Therefore, *Star* (hotel star class on a five-star scale, M = 2.982) and *MeanRating* (hotel overall average review rating on a scale of 1 to 5, M = 3.635) are used to measure hotels' managerial ability and service quality.

Furthermore, among all sampled hotels, 265 hotels belong to 23 different hotel chains, ranging from the luxury market to the economy segment. The difference between chain and independent hotels may contribute to a hotel's popularity. Their managerial strategies such as social media activity may also vary due to different business focuses and operations. A dummy variable *Chain* is created to indicate whether a hotel is a chain hotel (taking value of 1) or an independent hotel (taking value of 0). Moreover, the number of customers visiting a hotel may be a determinant of the number of reviews. The number of rooms is used as a proxy to measure the capacity of a hotel to accommodate visitors (*Size,* M = 108.222). Another point is that the total sum of reviews is also related to the duration of hotels on the review platform. This attribute is not included as a variable in the matching process and regression in this section, but this time factor is considered and reflected in the dependent variable of daily review volume.[7]

5.4.1 Model-free evidence

A two-sample t-test is conducted to determine if there is any significant difference in daily review volume between responding and non-responding hotels. Table 5-2 Panel A shows that responding hotels on average receive .452 online reviews per day (equivalent to 13.748 reviews per month and 164.976 reviews per year), while non-responding hotels have an average daily review number of .065 (equivalent to 1.977 per month and 23.725 per year). The difference between the two groups is statistically and practically significant ($p < .001$). Such significant difference exists between responding and non-responding hotels at each star class except for five-star hotels ($p = .137$). The model-free evidence suggests that review volume of responding hotels is significantly larger than non-responding hotels.

Table 5-2. T-test for review volume between responding and non-responding hotels

[7] Another reason to exclude *Age* in the matching process and regression in this section is that there is no significant difference between responding and non-responding hotels, as evident in the t-test results in Section 5.4.1.

Panel A: Review volume						
		$Response_{ihv-1}$=0		$Response_{ihv-1}$=1		t-test
Variable	Conditions	M	SD	M	SD	t-value
$DailyReviewVolume_h$		.065	.120	.452	.575	-11.250***
	Star = 0/1/1.5	.033	.117	.255	.287	-6.620***
	Star = 2/2.5	.065	.055	.242	.353	-4.037***
	Star = 3/3.5	.096	.127	.358	.534	-4.314***
	Star = 4/4.5	.079	.133	.600	.635	-3.656***
	Star = 5	.193	.263	.542	.612	-1.497

Panel B: Hotel attributes					
	Response = 0		Response = 1		t-test
Variables	M	SD	M	SD	t-value
$Star_h$	1.768	1.508	3.451	1.077	-19.903***
$MeanRating_h$	3.075	1.005	3.851	.709	-13.873***
$Chain_h$	.025	.155	.349	.477	-11.256***
$Size_h$	51	65.883	130.290	151.510	-8.527***
Age_h	3260.789	1342.259	3223.503	3110.440	.355

Note: *Response* is an indicator variable, demonstrating whether a hotel provides managerial responses. *DailyReviewVolume* is measured as the daily number of reviews in the duration of hotel presence on the site. *Star* is the hotel star class. *MeanRating* is the overall average customer review rating. *Chain* is a dummy variable indicating whether a hotel is a chain or an independent hotel. *Size* is measured by the number of hotel rooms. *Age* is the number of days from the date when a hotel receives its first review on the site to the end date of the sample period. * $p < .1$, ** $p < .05$, *** $p < .01$

It is too early to attribute this difference in review volume between responding and non-responding hotels to the provision of online managerial responses. Two issues are related to this problem. One is the endogeneity concern. Responding hotels may be inherently better managed and operated with quality service. As shown in Table 5-2 Panel B, the average star rating and customer review rating are higher for responding hotels, which also tend to be chain brands and to have larger capacity. These hotel-specific attributes could directly or indirectly contribute to the popularity of hotels in terms of offline guest visiting and online review writing, resulting in an increased or decreased number of online reviews. This leads to the second concern about the selection bias. Whether a hotel decides to engage in the online review platform is an outcome of self-selection. For example, hotels with better managerial capability are more likely to actively manage their social media presence and online reviews. Therefore, to measure the effect of responding on review volume, it is more accurate to compare review volume of responding and non-responding hotels that have similar characteristics.

5.4.2 Matching procedure and treatment effects

The provision of online managerial responses can be viewed as a 'treatment' that some hotels receive, and thus the research interest lies in the 'treatment effect' of responding on customer review volume. A control sample of comparable non-responding hotels is needed to examine how the review volume of the sampled responding hotels would be developed had they not provided managerial responses. PSM is adopted to create this control sample. Based on the likelihood that a hotel would be a responding hotel conditional on observable predictors, the PSM technique finds one or more 'similar' hotels (treatment non-recipient) for each responding hotel (treatment recipient), which alleviates the bias caused by the systematic difference between treatment recipient and non-recipient (Dehejia & Wahba, 2002). The treatment effect can be then analysed by comparing the treatment group and the control group (Rubin, 1973).

A binomial logistic regression specification is used to estimate the propensity score for each hotel, in which treatment (i.e., responding) is the dependent variable and predictors of responding are the independent variables. PSM outputs optimal results for a non-experimental study on treatment effects when appropriate predictors are used (Heckmann, Ichimura, & Todd, 1997; Wangenheim & Bayón, 2007). Considering that the probability of a hotel providing online responses is largely a result of self-selection, the four hotel characteristics—star class, average review rating, chain or independent brand, and size—are used to measure the likelihood of responding. Table 5-3 shows the result of logit regression. As expected, the identified four variables are strong predictors of the probability of response provision, suggesting that these factors need to be included in the regression estimation.

The next step is to match responding hotels to non-responding hotels using the computed propensity scores for each hotel. Various matching algorithms are available. It is started with k-nearest neighbour matching, in which treatment recipients are matched with non-recipients that have the closest propensity score. Both one-to-one and one-to-multiple (k is set to 4 as suggested in Abadie, Drukker, Herr, & Imbens, 2004) matching within common support region and without replacement are conducted, leading to a reduced sample of 798 hotels (525 in the treated group and 273 in the control group). A potential disadvantage of this approach is that the nearest match may actually be 'far away' (Caliendo & Kopeinig, 2008). To overcome this problem, a calliper matching is adopted by defining a radius (Dehejia & Wahba, 2002) with a width of .25 times the standard deviation of the propensity score (Rosenbaum & Rubin, 1985). The calculated maximum distance is about .07 and this is further restricted to .05. The nearest neighbour matchings with k equal to 1 and 4 are then repeated within the given calliper. The quality of matching using various algorithms is evaluated by calculating the percentage reduction in bias (PRB). It appears that allowing one match in the nearest neighbour is the most efficient. After matching, the standardised differences are less than 15%, and the PRB shows a substantial reduction in bias for the four predictors (see Table 5-3).

Table 5-3. Logit regression result and group means before and after matching

	Logit regression		Before matching		After matching			
Variables	Response		Control	Treated	Control	Treated	PRB	Std. Diff.
$Star_h$	.843***	(.076)	1.768	3.451	3.326	3.323	99.8%	-.2%
$MeanRating_h$	.322***	(.109)	3.075	3.851	3.753	3.748	99.3%	-.7%
$Chain_h$	2.749***	(.436)	.025	.349	.137	.112	92.4%	-7.0%
$Size_h$	.004***	(.001)	51	130.290	69.072	84.910	80.0%	13.6%***
_cons	-3.092***	(.369)						
N	1024		285	739	273	525		
Log-likelihood	-385.192							

Note: The matching results presented in this table are based on the k-nearest neighbour matching (k=1). *Response* = 0 in the control group and *Response* = 1 in the treated group. *PRB* is the Percentage Reduction in Bias. *Std. Diff.* is the standardised difference after nearest neighbour matching. Standard errors for the logit regression are in parentheses. * $p < .1$, ** $p < .05$, *** $p < .01$

Table 5-4 presents the treatment effect between the responding and comparable non-responding hotels. The results using different matching algorithms are very similar, and all are statistically significant. The k-nearest (k=1) and k-nearest within calliper (k=1, ε=.05) matching have the smallest mean bias after matching, and the one-to-one matching estimation is used as the base results. Before matching, the estimated difference in daily review volume between the two groups is .39. With the matched sample, the average treatment effect (ATE) is about .11, and the average treatment effect on treated (ATT) is about .14. This implies that responding hotels receive .14 more reviews per day (equivalent to additional 4.26 reviews per month and 51.10 reviews per year).

Table 5-4. Treatment effects using different matching algorithms

Group difference	k-nearest		Caliper	k-nearest within calliper		Kernel
	(k=1)	(k=4)	(ε=.05)	(k=1, ε=.05)	(k=4, ε=.05)	
Unmatched	.387***	.387***	.387***	.387***	.387***	.387***
	(.034)	(.034)	(.034)	(.034)	(.034)	(.034)
ATE	.114***	.125***	.136***	.114***	.125***	.134***
	(.027)	(.027)	(.026)	(.028)	(.028)	(.026)
ATT	.139***	.153***	.169***	.139***	.153***	.167***
	(.039)	(.039)	(.035)	(.039)	(.040)	(.035)
ATU	.066***	.072***	.073***	.066***	.071***	.072***
	(.020)	(.017)	(.014)	(.019)	(.017)	(.014)
On Support (treated)	525	525	525	525	525	525
On Support (control)	273	273	273	273	273	273
Mean Bias after matching	5.4%	10.7%	9.7%	5.4%	10.7%	10.2%

Note: The matching results presented in this table are the differences in daily review volume between the treated and control groups, using different matching algorithms. ATT is the Average Treatment Effect on the Treated. ATU is the Average Treatment Effect on the Untreated. ATE is the Average Treatment Effect. Bootstrap standard errors for ATT, ATU, and ATE are in parentheses. PRB is the Percentage Reduction in Bias. Standard errors in parentheses. * $p < .1$, ** $p < .05$, *** $p < .01$

PSM technique relies heavily on the selection of appropriate matching inputs. However, unobserved factors could affect whether hotels decide to provide online responses (treatment adoption) as well as the number of online reviews. The endogenous treatment effect of responding on review volume is further checked with linear estimations (see Table 5-5). Instead of daily review volume, the logarithm of review volume of each hotel in the sample period (*LnHotelReviewVolume*) is used as the dependent variable and the matching variables (i.e., *Star*, *MeanRating*, the logarithm of *Size*, and *Chain*) and the logarithm of *Age* are included as control variables. The first column in Table 5-5 shows maximum likelihood estimates indicating that providing response has a positive treatment effect and review volume of responding hotels is more than double the amount of non-responding hotels (if *Response*=1, review volume increases about 125%). The model is also estimated using a standard OLS regression that yields comparable results (see Table 5-5 Column 2).

Table 5-5. Estimations of endogenous treatment effects

$LnHotelReviewVolume_h$	(1) MLE	(2) OLS
$Response_h$	1.254***	1.139***
	(.119)	(.113)
$Star_h$	.406***	.421***
	(.040)	(.043)
$MeanRating_h$	-.010	-.004
	(.062)	(.062)
$LnSize_h$	.443***	.448***
	(.050)	(.050)
$LnAge_h$	.596***	.594***
	(.040)	(.041)
$Chain_h$	.429***	.451***
	(.080)	(.080)
_cons	-2.982***	-2.980***
	(.411)	(.413)
N	1003	1003
Log-likelihood	-1855.816	-1482.132
R^2		.632
Chi-square	1615.941	
p	.000	.000

Note: The first column presents the maximum likelihood estimations of treatment effect. The logit regression of *Response* shows that all predictors of *Response* are significantly positive at p < .05, and these results are not presented in this table. Standard errors in parentheses. * $p < .1$, ** $p < .05$, *** $p < .01$.

5.5 Effects of Response Visibility: Responding Activeness

5.5.1 Methods and results

The previous section analyses whether the provision of managerial response has an impact on the volume of customer reviews. A more interesting question is how the responding styles can affect such impact in the long run after hotels start to respond. Specifically, this section investigates the visibility of online managerial response and how its visibility contributes to the difference in review volume between responding and non-responding hotels. To test Hypotheses 2a and 2b, a subsample is created only keeping review observations of responding hotels after the date of each responding hotel's first managerial response. The subsample includes 692 hotels, 642,501 customer reviews, and 358,752 managerial responses from March 2004 to February

2016. The effects of response visibility are investigated using panel data with the dependent and explanatory variables organised at hotel–month level.

The dependent variable is the logarithm of a hotel's number of reviews in a calendar month (*LnReviewVolume*), and the explanatory variables are the two variables of response visibility (*ResponseRatio* and *LnResponseDays*) in the previous month. Besides, a vector of control variables is included in the model to address a few endogeneity concerns. First, as discussed in the previous section, hotels providing quality services may naturally attract more consumers to stay, potentially leading to a higher volume of reviews, and these hotels are more likely to provide managerial responses in a professional way, leading to a higher level of response visibility. Therefore, the three hotel-specific factors—*Star, LnSize,* and *Chain*—are included to account for hotel-level heterogeneity.[8] Next, apart from the possible influence of offline popularity, online popularity—the cumulative number of reviews for a hotel at a time point—may impose crowd effects on potential reviewers' behaviour. It may also affect managers' response behaviour, given the varying volume of reviews they receive. To tackle this issue, a variable *LnCumReviewVolume* is included—that is the logarithm of the cumulative review volume of a hotel in the previous *t-1* periods. Moreover, some hotel characteristics change in the period that may have influence, such as customers' ratings and the duration of a hotel's presence on the online review platform since the date of its first review. Given that the customer rating is a reflection of customer satisfaction and service quality, the moving average of review rating (*CumAverageRating*) of a hotel in the previous *t-1* periods is used to control for unobserved time-varying managerial expertise and exogenous shocks. To control for time effect, the cumulative number of days (*LnAge*) of a hotel's appearance on the site until the end of the *t-1* period is included. A time dummy for each month is also included to account for time trend of review and response behaviour that is common to all hotels. The Pearson correlation coefficients in Table 5-6 show that variables are significantly correlated at $p < .05$ level. Multicollinearity is not a concern for the variables of interest given the low VIFs compared to the common threshold. The model is specified as:

[8] The value of these variables is at a fixed time point (i.e., the time point of data collection), but these attributes are not strictly time-invariant (e.g., star class upgrade/downgrade, size expansion, brand acquisition etc.). The purpose of including these variables in the model is to account for differences among sampled hotels in these aspects.

$$LnReviewVolume_{ht} = \beta_1 ResponseRatio_{ht-1} + \beta_2 LnResponseDays_{ht-1} + \beta_3 Control_h + \varepsilon_{it}$$

where the dependent variable and the two response variables are at monthly level, and *ResponseRatio* and *LnResponseDays* are one month lagged; $Control_{ht-1}$ is a vector of control variables, including time-varying hotel factors—*LnCumReviewVolume*, *CumAverageRating, LnAge* (monthly level, one month lagged), time-invariant hotel factors—*Star, LnSize,* and *Chain,* and time dummies.

Table 5-6. Summary statistics, correlation matrix and multicollinearity check

	1	2	3	4	5	6	7	8	9	M	SD	VIF
1.LnReviewVolume$_{ht}$	1									2.281	1.128	
2.ResponseRatio$_{ht-1}$	.299*	1								.452	.421	1.050
3.LnResponseDays$_{ht-1}$	-.263*	.080*	1							2.071	1.149	1.090
4.LnCumReviewVolume$_{ht-1}$	.704*	.262*	-.242*	1						5.297	1.466	4.540
5.CumAverageRating$_{ht-1}$	.332*	.217*	-.123*	.236*	1					3.868	.659	1.430
6.Star$_h$	.226*	.247*	-.045*	.170*	.449*	1				3.634	.904	1.520
7.LnSize$_h$	.607*	.222*	-.157*	.398*	.168*	.326*	1			4.518	.976	2.010
8.LnAge$_h$	.224*	.115*	-.115*	.748*	-.010*	.043*	-.015*	1		6.617	1.099	3.580
9.Chain$_h$	.293*	.163*	-.098*	.148*	.113*	-.074*	.417*	-.069*	1	.346	.476	1.310

Notes: Variables 2–5 are one month lagged. * $p < .05$

Considering that the data on monthly review and response is clustered at the hotel level, a multilevel model (two-level model) is adopted, with the review/response at the first level and the hotel as the second level indicator. The random effects model estimates the group effects and group level predictors at the same time. Furthermore, there might be time effects and such time effects may vary across individual hotels. Hence, month dummies are added to the model and also the time factor (i.e., month) is included at the group level to allow for random slopes across different hotels. Table 5-7 Columns 1-3 show the results. As expected, response days are negatively associated with review volume (β = -.028, $p < .001$). For example, a 10% decrease in the monthly average time intervals (days) between reviews and the associated responses from the service provider leads to about .28% increase in the review volume in the following month. This suggests that response speed has a positive impact on review volume, which supports hypothesis 2b, although the effect size is small. The effect of response ratio has statistical significance when only response ratio is considered (β = .080, $p < .001$); however, this attribute becomes insignificant when response speed is included (β = .015, $p = .377$). These findings together support the hypothesis about the positive effect of response visibility on future review volume, with response speed making a more substantial contribution.

Table 5-7. Effects of response visibility

LnReviewVolume$_{ht}$	Multilevel model			OLS		
	(1)	(2)	(3)	(4)	(5)	(6)
Fixed effects						
ResponseRatio$_{ht-1}$	.080***		.015	.118***		.009
	(.015)		(.017)	(.018)		(.019)
LnResponseDays$_{ht-1}$		-.028***	-.028***		-.026***	-.026***
		(.004)	(.004)		(.005)	(.005)
LnCumReviewVolume$_{ht-1}$	.379***	.529***	.532***	.535***	.464***	.465***
	(.021)	(.022)	(.022)	(.028)	(.033)	(.033)
CumAverageRating$_{ht-1}$	.182***	.156***	.155***	.241***	.179***	.178***
	(.027)	(.026)	(.026)	(.033)	(.039)	(.039)
Hotel-level predictors	Yes	Yes	Yes	No	No	No
Hotel dummy	No	No	No	Yes	Yes	Yes
Time dummy	Yes	Yes	Yes	Yes	Yes	Yes
Intercept	-1.187***	.366**	.375**	.922**	1.627***	1.626***
	(.385)	(.184)	(.184)	(.360)	(.230)	(.229)
Random effect variances						
Hotel level	.329***	.076***	.076***			
	(.064)	(.020)	(.020)			
Month time effect	.000***	.000***	.000			
	(.000)	(.000)	(.000)			
Monthly review/response level	.206***	.167***	.167***			
	(.005)	(.005)	(.005)			
N	35068	22854	22854	35068	22854	22854
R-square Adjusted				.815	.833	.833
Log-likelihood	-	-	-	-23231.4	-	-
	23953.12	13043.01	13042.19		11943.46	11943.15

Note: The independent variables of response and review are one month lagged. The multilevel models present maximum likelihood estimations. In the OLS models, the time-invariant group-level predictors (i.e., *Star, LnSize,* and *Chain*) are omitted. All estimations have robust error terms clustered at the hotel level. Standard errors in parentheses. * $p < .1$, ** $p < .05$, *** $p < .01$.

In addition, the results show that review volume is largely affected by former reviewers in terms of whether they write reviews and how they rate the service. As Table 5-7 shows, the cumulative number of reviews (β = .532, $p < .001$) and moving average of customer ratings (β = .155, $p < .001$) of hotels established online in the previous periods are positively related to future review volume. For instance, an approximate 10% increase in the aggregated review volume or a 1 score increase in the moving average of ratings may lead to about 5.32% or 15.5% increase in the next period's review volume. This implies the significance of the crowd effect on customer engagement behaviour. Furthermore, with regard to the random effects, 3.13% of the variance in review volume is attributed to the differences among hotels and the time effect is trivial, while the differences in the attributes within each hotel account for the majority of the effects.

5.5.2 Robustness checks

Several additional tests are conducted to check the robustness of the results. First, the model is re-estimated using a pooled OLS regression with hotel dummies and time dummies. The estimations (Table 5-7 Columns 4-6) are very similar to the baseline results. Consistent with the multilevel random effect estimations, the response days are negatively related to review volume (β = -.026, $p < .001$), suggesting a positive effect of response speed on future review volume. Response ratio still presents statistical insignificance (p = .573). Meanwhile, the positive effects of two review descriptors remain significant. Furthermore, the results may be sensitive to potential outliers. The variables of response volume and response are winsorised at the 0th and 99th percentiles and the results are still robust (not reported in the table).

Second, it is worth pointing out that the time dummies are significant after the year 2009. This is in line with the growing trend in reviewing and responding behaviour starting from that date. To further check the robustness, all data before the year 2009 that account for 4% of the hotel–month observations is eliminated. The remaining data in the period 2009–2016 is used to re-estimate the multilevel model. The estimations (Table 5-8 Panel A) confirm the main results that response speed is positively associated with review volume (β = -.028, $p < .001$) and response ratio has no statistically significant influence (β = .010, p = .573).

Table 5-8. Robustness checks

$LnReviewVolume_{ht}$	Panel A: After the year 2009			Panel B: One-time reviewer			Panel C: Response volume		
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
Fixed effects									
$ResponseRatio_{ht-1}$	$.072^{***}$		.010	$.076^{***}$		.012	$.075^{***}$		$.075^{***}$
	(.015)		(.017)	(.015)		(.017)	(.006)		(.006)
$LnResponseDays_{ht-1}$		$-.028^{***}$	$-.028^{***}$		$-.029^{***}$	$-.029^{***}$		$-.028^{***}$	$-.029^{***}$
		(.004)	(.004)		(.004)	(.004)		(.004)	(.004)
$LnCumReviewVolume_{ht-1}$	$.344^{***}$	$.521^{***}$	$.523^{***}$	$.375^{***}$	$.532^{***}$	$.534^{***}$	$.499^{***}$	$.529^{***}$	$.500^{***}$
	(.022)	(.023)	(.023)	(.021)	(.021)	(.021)	(.021)	(.022)	(.021)
$CumAverageRating_{ht-1}$	$.173^{***}$	$.162^{***}$	$.162^{***}$	$.188^{***}$	$.156^{***}$	$.155^{***}$	$.149^{***}$	$.156^{***}$	$.144^{***}$
	(.029)	(.028)	(.028)	(.027)	(.025)	(.025)	(.024)	(.026)	(.024)
Intercept	.036	$.608^{**}$	$.614^{**}$	-1.181^{***}	$.432^{**}$	$.439^{**}$	$.302^{*}$	$.366^{**}$	$.409^{**}$
	(.235)	(.242)	(.243)	(.382)	(.178)	(.179)	(.172)	(.184)	(.172)
Random effect variances									
Hotel level	$.411^{***}$	$.087^{***}$	$.087^{***}$	$.306^{***}$	$.069^{***}$	$.068^{***}$	$.064^{***}$	$.076^{***}$	$.061^{***}$
	(.008)	(.031)	(.031)	(.062)	(.016)	(.016)	(.015)	(.020)	(.014)
Month time effect	$.000^{***}$	$.000^{***}$	$.000^{***}$	$.000^{***}$	$.000^{***}$	$.000^{***}$	$.000^{***}$	$.000^{***}$	$.000^{***}$
	(.000)	(.000)	(.000)	(.000)	(.000)	(.000)	(.000)	(.000)	(.000)
Monthly review/response level	$.200^{***}$	$.165^{***}$	$.165^{***}$	$.209^{***}$	$.172^{***}$	$.172^{***}$	$.166^{***}$	$.167^{***}$	$.166^{***}$
	(.005)	(.005)	(.005)	(.005)	(.005)	(.005)	(.005)	(.005)	(.005)
N	33905	22671	22671	34896	22742	22742	22857	22854	22854
Log-likelihood	-22823.359	-12854.748	-12854.418	-24085.775	-13292.720	-13292.186	-12920.306	-13043.014	-12868.756

Note: The independent variables of response and review are one month lagged. Hotel-level predictors (i.e., *Star, LnSize, LnAge* and *Chain*) and month dummies are included. The multilevel models present maximum likelihood estimations. All estimations have robust error terms clustered at the hotel level. Standard errors in parentheses. * $p < .1$, ** $p < .05$, *** $p < .01$.

Next, the data includes some repeat reviewers who have multiple reviews for the same hotel. Godes and Mayzlin (2009) demonstrate that the effects of firm-generated messages in the word-of-mouth marketing campaign vary with the degree of customer loyalty. Their findings suggest that exogenous word-of-mouth created by firms is more impactful and raises awareness among less loyal customers because they are less informed than loyal customers, who have already formed strong ties and opinions about the firm. Gu and Ye (2014) also hint that there might be a self-selection issue among returning customers who are more likely to write reviews. The information distortion derived from individual preference may affect the decision-making process (Chaxel & Han, 2018). Therefore, the reviews written by returning customers (i.e., a customer writes more than two reviews of the same hotel in the sample period) are excluded to check the sensitivity of results to customers' heterogeneous preference. Panel B of Table 5-8 shows that the results are robust to measuring one-time reviewers. The response speed remains significant (β = -.029, p < .001), and response ratio has no significant impact (β = .012, p = .469). The magnitude of coefficients is very similar to the base results, but the random effects of monthly reviews and responses are slightly stronger when only one-time reviewers are considered.

Finally, in Panel C of Table 5-8, instead of response ratio, the actual number of responses of a hotel in each calendar month is used as a proxy for response visibility. The results show that the response days retain a significant and negative relationship with the review volume in the following period and the magnitude is very similar to the base results (β = -.029, p < .001). However, the response volume is also significant (β = .075, p < .001). One possible explanation is that the absolute value of response volume cannot reflect the number of reviews that a hotel needs to respond to in a period (Xie et al., 2016). A higher number of responses and a higher response ratio are not equal, which may bias the results. The base number of customer reviews should be taken into consideration in evaluating the effect of response volume. Nevertheless, these results support the importance of keeping active to increase the visibility of firms' communications with potential reviewers.

5.6 Discussions and Conclusions

5.6.1 Summary of findings and theoretical implications

This study examines online firm and customer engagement issue by studying the treatment effect and long-term behavioural effect of managerial responses to customer reviews. The key findings are summarised and discussed as follows. First, using review and response data from a single online review platform, a PSM matching based on the hotel service quality and managerial expertise is conducted to compare the review volume between responding and non-responding hotels. It is found that the daily and total numbers of customer reviews of responding hotels are larger than those of comparable non-responding hotels. This finding is in line with findings of prior studies (e.g., Ye et al., 2010; Proserpio & Zervas, 2017; Chevalier et al., 2017) that providing online managerial responses has a positive impact on customer review volume. This implies that responding hotels present greater online popularity compared to their counterparts in the market.

Furthermore, after responding hotels implement the online response strategy, the visibility of managerial responses on the site plays a vital role in the longer term to affect customers' participation in reviewing. A panel regression approach is adopted to test two response visibility indicators, response ratio and response speed. Controlling for possible hotel heterogeneity and time effects, the results show that response speed has a strong positive effect on future review volume. This finding echoes the significance of timing in the service recovery literature (e.g., Davidow, 2003; Homburg & Fürst, 2007; Sparks et al., 2016). The novel aspect of this finding is that the importance of promptness is emphasised in relationship to the visibility of managerial messages in online communications with potential reviewers rather than the effects of response promptness on existing low-satisfaction consumers.

In addition, in contrast to the speculation about the effect of response ratio, the empirical results do not show a significant influence of response ratio on future review volume. This finding contradicts that of Xie et al. (2016), which documents a strong positive association. One possible explanation is that this study considers the two response descriptors at the same time. These two variables are correlated, and both affect review volume. It is evident in the results that response ratio is significant when it is the only response variable included in the model, but it becomes insignificant when the variable of response speed is included. This implies that response ratio is trivial and less decisive, while response speed plays a more vital role in increasing the response visibility on the site and thus affecting potential reviewers. An exogenous factor that affects the visibility of responses is the update speed of customer reviews, i.e., how many reviews a business receives in a certain period. Quicker responses increase the possibility of responses being displayed on the first few pages and thus being easier for customers and potential reviewers to see. However, a higher response ratio does not necessarily indicate the responses are prompt. If the response frequency cannot compete with the review update speed, even if a large proportion of reviews are responded to, it is hard for the managerial responses to be posted on the first few pages, leading to a reduced power in influencing the propensity for customer engagement.

In addition, the findings demonstrate that information about previous reviews has a substantial influence on potential reviewers. The average rating and the aggregated number of reviews are positively related to the review volume in the subsequent period. This suggests that the reviewing decision is largely affected by how other reviewers engage and rate the service. This presents a crowd effect in the online engagement behaviour, which further accentuates the need for effectively managing the popularity created by the online community.

Altogether, these findings suggest that customers' engagement intention and behaviour are influenced by firms' engagement in the online conversations. This contributes to the engagement literature (e.g., Eisingerich et al., 2015; Mathwick & Mosteller, 2017; Pansari & Kumar, 2017; Van Doorn et al., 2010) by determining that firm engagement is a motivational driver for customer engagement behaviour. Apart from self-motivation for word-of-mouth sharing (Berger, 2014), there is a spillover effect of the managerial intervention on reviewing behaviour of the community members. A business being active on the social media can facilitate interactions between customers and firms, which can attract, encourage and stimulate online users, especially potential reviewers, to engage in online reviewing and communications. This study also contributes to the management research in relation to social media efforts by investigating the effect of online firm-generated messages that has been understudied in the current literature (Harmeling et al., 2017; Kumar et al., 2016). Studying the behavioural effects of response provision and response visibility suggests that firms' strategic participation in online communications can potentially create leading influence and draw wider attention, which makes it an effective tool to enhance online popularity and social influence.

5.6.2 Implications for practice

These discussions clearly demonstrate that firms' active presence in social media can stimulate customer engagement behaviour. The business' strategic and voluntary exposure on online social sites can help gain customers' attention, expand the consumer network, manage customers, and enhance social influence and online popularity, all potentially leading to favourable outcomes. This requires firms to make strategic changes with "committing to long-term paths or trajectories of competence development" (Teece et al., 1997, p. 529). For firms that have not established an online presence in the network, providing managerial responses would be an option to kick-start engagement in online firm-customer communications and active management of their social media presence. For firms that have adopted social media to implement marketing activities, it is important to keep the engagement and communication as a consistent practice. Businesses should respond in a faster and frequent way to make sure the managerial effort is manifest to customers. When there is a large number of reviews, firms could respond to the most recent reviews first to increase the visibility of responses on site and thus influence review readers and potential review writers. The visibility of firms' presence and the activeness of online engagement are expected to have a continuous positive impact in the longer term. This creates strategic value for managing customers and potentially for financial outcomes.

5.6.3 Limitations and future research

A few limitations should be acknowledged. First, this study focuses on online popularity as demonstrated by the number of customer reviews on a review website. It does not consider offline popularity, such as the actual number of visitors, and its potential influence on the review volume. Future research may extend this study by examining the relationship between offline and online popularity and the possible impact of managerial response on sales/revenue generation. Second, the included control variables of hotel characteristics are not exhaustive. Additional variables, such as price and location, can be added to the model to control for offline popularity. Third, the research setting to investigate the business social media presence and activeness in this study is an online community-based review platform. This is a third-party organised communication channel, which may present some policy-related issues that impede or affect how firms engage. Furthermore, the review-response communication is less firm-initiated. It would be interesting to examine the interplay between firm engagement and customer engagement behaviour by using "firm-initiated marketing communication in its official social media pages" (Kumar et al., 2016, p. 7), given that the corporate resources allocated to managing the channels are different (Ashley & Tuten, 2015).

CHAPTER 6

CONCLUSIONS

The final chapter reflects on the issue of firm engagement in online social interactions (OSIs) examined in this doctoral research. The chapter starts with a summary of the research findings by revisiting the research questions set out at the beginning. This is followed by a discussion on the theoretical contributions to knowledge, implications for managerial practices, and suggestions for future research. Figure 6-1 outlines the key findings from each study and their practical implications and theoretical contributions, and highlights how the doctoral research collectively contributes to the knowledge of the field and the direction in which future research can be extended. The details of these elements will be further elaborated in the following sections.

6.1 Research Findings

The thesis examines the effectiveness of firm engagement in OSIs and develops strategies for firms to manage voices from the online crowd of consumers effectively so as to create business benefits. Built upon prior scholarly work (Godes et al., 2005; Zhang et al., 2011), firm engagement in this research is defined as a firm's strategic intention and behavioural manifestation, motivated by multifaceted incentives, of engaging in the OSI network and acting on OSIs among consumers. Such engagement is manifested in firm–customer interactions via digital platforms, on which firms play three different but non-mutually exclusive roles. A conceptual framework is proposed to provide a holistic view of firm engagement in OSIs by identifying the observer, participant, and strategic leader roles in the OSI network and clarifying activities and capabilities associated with these different roles. This finding answers the first research question of the business roles in the OSI network.

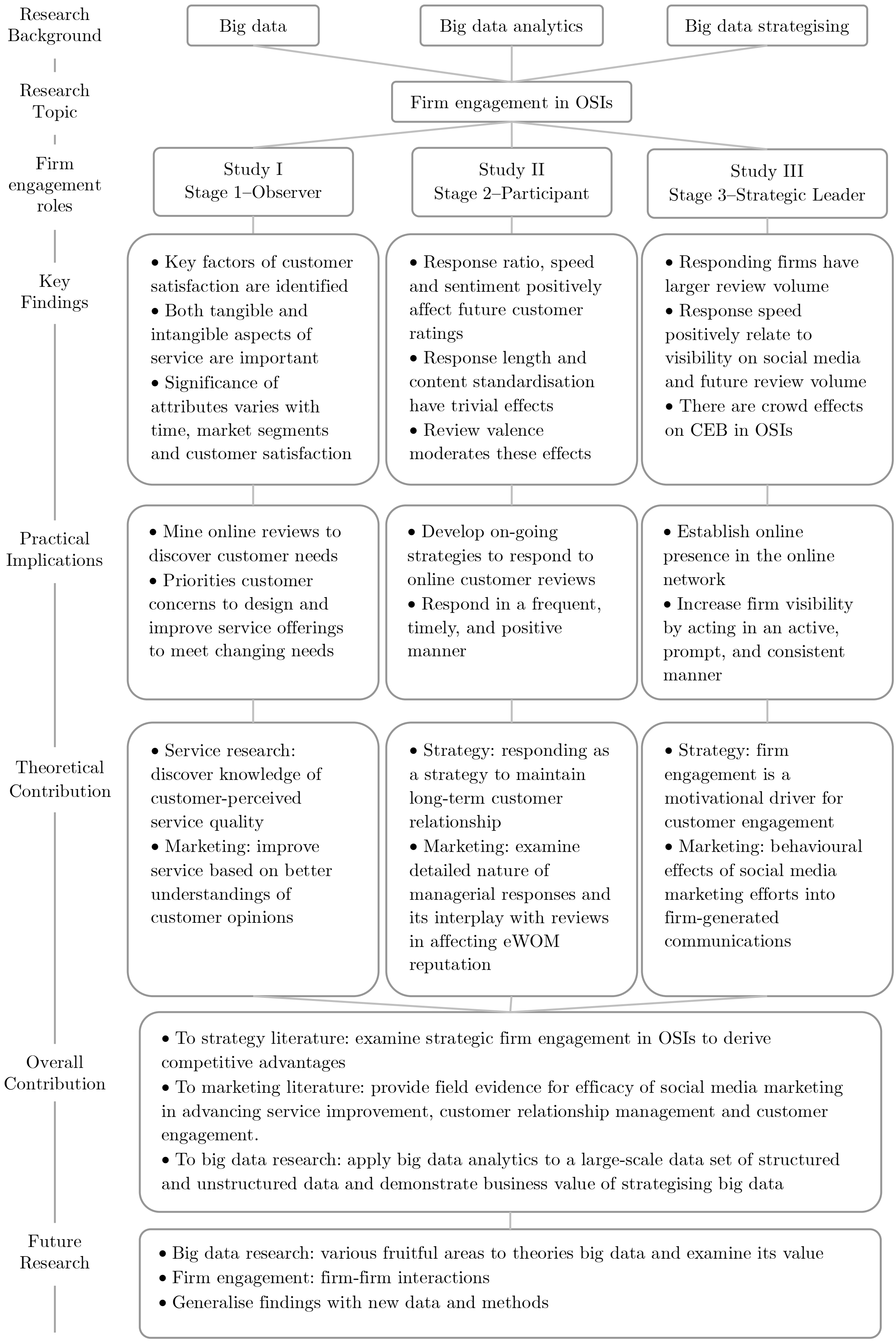


Figure 6-1. Research summary

In response to the second and third research questions—"What are the business impacts of firms managing and engaging in OSIs?" and "To what extent can firm engagement in OSIs create benefits for businesses?"—three empirical studies have been conducted to examine effects of the identified firm engagement roles on business performance. Specifically, the three studies demonstrate the relationship between firm engagement and business performance through analysing the consequences of firms learning, participating and leading OSIs, and they also examine the significance of the potential business impacts and discuss strategies for firms to interact with customers so as to exploit OSIs in creating strategic values. Key findings from the three studies are summarised as follows.

Chapter 3 investigates a strategy for firms to learn about consumers' opinions and to discover knowledge from consumers' online comments for a service improvement purpose. This study contends that online customer reviews can be a rich source of information, and mining customer-generated reviews can equip organisations for identifying dimensions of consumer-perceived service quality and prioritising improvement areas. The strategy for discovering knowledge from consumers' online comments is demonstrated by applying a Latent Dirichlet Allocation (LDA) topic-modelling approach to analyse the large collection of review text. This study discovers the predominant attributes that affect consumers' perception of hotel services and illustrates their importance using the SERVQUAL construct. One novel finding of the topic-modelling analysis is that a few important attributes such as guest review and homeliness emerge from the text-mining analysis which have been overlooked by prior studies, while some attributes such as housekeeping and room service highlighted by the current literature have not appeared in the statistical model results. This adds to our knowledge by identifying the detailed elements and shifting consumer attention towards service quality attributes other than those well-known or generic dimensions such as cleanliness and quietness (e.g., Rhee & Yang, 2015; Banerjee & Chua, 2016). Moreover, the significance of consumer-perceived service quality changes along the timeline and depends on the market segments that firms are positioned in and varies with the level of customer satisfaction. Compared to prior research (e.g., Guo et al., 2017), these findings from incorporating the longitudinal analysis, sentiment analysis and segment analysis in the topic-modelling approach provide new insights into consumers' changing behaviour over time, different opinions of satisfied and unsatisfied customers, and diverse expectations of service standards for different market segments. With a better understanding of the service dimensions that consumers value most, firms will be able to know where to target improvements of service offerings. This study highlights the potential for service organisations to engage in OSIs and improve service quality by capitalising on the knowledge derived from user-generated content (UGC).

Chapter 4 explores strategies for firms to directly participate in OSI activities. This study examines online managerial responses to customer reviews, assessing the efficacy of responding, in its various styles, in terms of future customer ratings on the review platform. This is a deviation from prior research that focuses on the changes in ratings before and after the adoption of a managerial response policy (e.g., Proserpio & Zervas, 2017; Wang & Chaudhry, 2018), with this study directing attention towards the mechanism that yields such outcomes in the long term. By examining the response approaches after responding firms implement the policy, results show that a higher response ratio and speed can lead to higher ratings in the subsequent period. In addition, a novel aspect of this study is that text mining and sentiment analysis are applied to investigate the detailed nature of managerial responses not explored by previous research (e.g., Gu & Ye, 2014; Xie et al., 2016), which are communication styles embedded in response text—namely, response length, content standardisation and sentiment. It is found that the tone of a response plays a role, meaning that positive sentiment is helpful in improving future ratings, while the number of words and similarity in response content have a trivial impact. These findings underline how non-verbal cues hidden in managerial responses influence ratings posted by prospective reviewers. In addition, considering the interplay between responses and reviews, this study tests their joint effects and finds that the effects of managerial responses are moderated by review valence. This finding helps develop solutions for dealing with different types of reviews rather than being limited to the complaint scenario (e.g., Min et al., 2015; Sparks et al., 2016). Overall, this study presents field evidence on the efficacy of managerial communications with customers on online platforms in consolidating favourable electronic word-of-mouth (eWOM) and moderating effects of bad words. It highlights the responsiveness of firms in participating in online communications so as to enhance customer relationships and eWOM reputation.

Chapter 5 examines the social influence of firms' strategic path of being active and visible in the OSI network. Specifically, firms' active management of eWOM within social media is a strategic decision which may affect customers' engagement behaviour when participating in online communications. This conclusion is derived from two analyses of firms' presence and activeness in an online review community. The first analysis detects the relationship between firms' responding and the review volume, and the results show that responding firms have a greater number of customer reviews compared to non-responding firms. This finding illustrates a firm-leading influence on customer word-of-mouth behaviour, apart from customers' self-motivation and individual determinants (e.g., Hennig-Thurau et al., 2004; Van Doorn et al., 2010), by identifying the positive impact that firms have on stimulating potential reviewers by being present in online communications. The second analysis reveals that the visibility of managerial responses is critical in effecting firms' social media management efforts in the long run. The results show that the speed of responding has a positive influence on future review volume, while response ratio plays a trivial role. It implies that on a long-term basis, visibility of firm-generated online messages determined by response speed positively affect customer engagement behaviour. Taken together, the third study supplements the research on firm-generated content through examining its behavioural value in stimulating customers' social media engagement (e.g., Kumar et al., 2016). The derived insights suggest that businesses manage social media and communicate with consumers in an active and prompt way so as to create strategic benefits of expanding the online network and enhancing online popularity.

On the whole, this thesis demonstrates competitive advantages stemming from firm engagement in OSIs. The discovered favourability leads to a discussion regarding the last research question of strategic management of OSIs. With a conceptual extension and empirical investigation of an integrated view of firms managing OSIs as constructed in Godes et al. (2005), this research provides a holistic and empirical examination of firm-level strategies for engaging in OSIs. Diverging from past studies that concentrate on customer-to-customer interactions in the OSI network (King et al., 2014), firm engagement in this research is validated as firm–customer interactions in the online participative and social environment. Such interactions between firms and customers can include mining information from customer reviews, which offers insights into collective opinions about service quality and can serve as a knowledge base for businesses to spot market opportunities and tailor efforts to improve customer experience (Chapter 3). Firm–customer interactions can also include establishing conversations with consumers by responding to reviewers' comments, which influences both prospective consumers' perception of firm reliability and future ratings (Chapter 4). Firm engagement can include further proactive action plans in which businesses initiate and promote social media activities with business exposure and high visibility, and this is proven to be powerful in stimulating customer engagement behaviour in eWOM (Chapter 5). These proposed strategies demonstrate that the managerial efforts devoted to engaging in OSIs lead to enhanced effectiveness and efficiency in online marketing communications and business operations. This accentuates the fundamental necessity of business transition to data-driven decision-making and commitment to engage in the new digital landscape.

6.2 Theoretical Contributions

This doctoral research was motivated by the insufficient understanding of big data value in business and in particular how firms strategise customer-generated data through engagement in the OSI network. In exploring answers to the research questions, three studies are carried out to examine firm engagement in OSIs. While each study presents novel perspectives to understand firm engagement role and impact (as discussed in each chapter), the thesis as a whole makes several theoretical contributions. Table 6-1 summarises the contributions of each study in response to the research questions as well as how the three studies together make contributions to the strategy, marketing, and big data research.

Table 6-1. A summary of research contributions

Firm engagement	Study I	Study II	Study III
RQ1: firm roles?	Stage 1 - Observer	Stage 2 – Participant	Stage 3 – Strategic Leader
RQ2: business impact? RQ3: to what extent? RQ4: strategic plan?	Discover customer concerns about service quality by mining user-generated content • What consumers value most in relation to service quality? • How the dimensions' importance change over time, across market segments and vary with satisfaction?	Impact of managerial responses on future customer ratings • What are the communication styles embedded in managerial responses? • How and to what extent the response action and communication styles affect future ratings?	Behavioural value of firms' social presence and activeness in stimulating customer engagement behaviour • Whether and to what extent firm responding affect review volume? • To what extent the visibility of managerial responses affects review volume?
Contributions	• Identify service quality attributes directly from user-generated reviews • Track the changes in dimensions' significance overtime, across market segments and customer satisfaction • Synthesis quality dimensions and define priorities for decision-makings	• Discover the on-going impact of responses on future ratings in the long run after policy adoption • Identify important communication styles of responding that have an impact on future ratings • Take into account the interplay between review and response and their joint effect on future ratings	• Identify firm-customer interaction as a motivational driver for customer engagement in eWOM • Firm's online presence has behavioural value • Maintain online activeness to enhance visibility is beneficial to stimulating customer engagement behaviour
	• Strategy literature	• Advance understanding of OSIs and social media strategies • Provide a holistic view of firm engagement in OSIs • Add evidence to the efficacy of firm engagement	
	• Marketing literature	• Digital marketing (social media marketing): manage UGC and online interactive media • Emerging research on online firm–customer interaction and firm-generated content • Dynamic marketing capabilities: leverage UGC to understand consumers, enhance customer relationship management and eWOM engagement with effective communications	
	• Big data research	• Clarify the research frontier of big data • Combine different analytics techniques to a large-scale real-application dataset	

First of all, this research advances our understanding of OSIs and social media strategies. As identified in the literature review chapter (Sections 2.4.6 and 2.5.4), there is a proliferation of UGC and a growing scholarly interest in social media management; nevertheless, few studies have explored how firms should act on the voices from the big crowd of customers. Over the years, while extensive studies have been exploring the antecedents and consequences of customers' OSI behaviour (e.g., De Matos & Rossi, 2008), the implications of OSIs for strategic management remain under-explored. Among a handful of research studies on business management of OSIs, firm engagement behaviour has been explained in isolation, which results in a lack of systematic vision. The thesis discusses firm-level engagement strategies for managing OSIs. The investigation into three stages of firm engagement in an integrated framework develops the current research on OSI engagement (e.g., Van Doorn et al., 2010; Mathwick & Mosteller, 2017) by presenting a comprehensive view of OSI behaviour with firms involved in the OSI network. The focus on business actions extends OSI management research (e.g., Godes et al., 2005) by synthesising and distinguishing the activities and capacity of various business roles in the OSI network as well as empirically examining the efficacy of business actions of observing, participating and leading OSIs among customers. More broadly, this thesis contributes to the research on social media strategy (e.g., Kaplan & Haenlein, 2010) by developing and validating firm engagement strategies to manage OSIs so as to gain business benefits and competitive advantages in the marketplace.

A second contribution of this research is derived from the first—to present evidence and explain the capability of different strategies for firm engagement in OSIs in order to enhance social media marketing. In spite of the fact that social media is becoming a desired and efficient channel connecting consumers and marketers, UGC remains an under-utilised source of information, and research on firms' strategic use of social networking sites and UGC remains in its infancy (Goh et al., 2013). The current research explores firm–customer interactions on an online review site. It contributes to the broad digital marketing literature (e.g., Lamberton & Stephen, 2016; Parsons, Zeisser, & Waitman, 1998) with detailed discussions and examinations of leveraging UGC and firm-generated marketing communications via interactive media. In particular, findings from this research support the model of dynamic marketing capabilities (Barrales-Molina et al., 2014). It is proved that accessing market knowledge through mining and analysing information provided by customers would not only help service organisations gain a better understanding of market desires, but also inform business actions on service improvement, leading to enhanced performance and customer satisfaction. In addition, evidence of improved customer relationship management is documented in this research by exploring efficient communication styles of interacting with customers and encouraging customer engagement in OSIs. This adds weight to the idea of dynamic marketing capability by emphasising the integration of social media and digital marketing strategies so as to increase firms' customer relationship management capabilities (Wang & Kim, 2017) and ability to reconfigure business competences.

An additional major contribution of this doctoral research stems from the examination of the value and strategic use of big data. This research is built on a comprehensive review of big data literature, with a focus on the business and management domain. It clarifies the current research frontier of big data and depicts a clearer path towards management progress in big data value achievement. To illustrate the potential for strategising big data, this research adopts a data-driven approach aimed at empirically analysing massive amounts and various types of data. Several advanced analytics techniques, such as text mining and sentiment analysis, are employed to quantify unstructured human conversations through mining natural language, thus discovering thematic structure and semantic information. In combination with other quantitative data on instruments, the quantified textual data is used to construct statistical models and predict correlations. Compared to previous research using experiment, interview or conceptual methods, the mixed methods design and diagnostics and/or predictive analysis of massive online publically available OSI data capture specifically detailed aspects of customer and business activities and provide a comprehensive analysis of the research problem. This contributes to big data research in management, specifically in marketing and strategic management (e.g., Erevelles, Fukawa, & Swayne, 2016; Wedel & Kannan, 2016), in terms of enriching our knowledge of data-driven strategies as well as generalising this approach to enable concrete interpretation of real-world situations in the ever-changing business environment.

6.3 Practical Implications

How can firms act on the insights developed in this research and harvest the business value of online customer voice more effectively? Several practical implications are identified. First, customer opinions expressed publicly on online social platforms are valuable information resources to embrace in developing and renewing dynamic marketing capabilities. Firms are advised to enhance business analytics capacity to collect user-generated online messages. This available data can be examined on a large scale to reveal hidden information, unknown patterns and collective opinions about brands, products, and services. Companies may benefit from such an approach of discovering exact information and details of the market so as to make accurate judgements. For service organisations in particular, using big consumer data to improve service design and delivery is a priority in co-creating value with customers to respond to the rapidly changing market (Ostrom, Parasuraman, Bowen, Patricio, & Voss, 2015). Dedicated firms might listen to and crowdsource consumers' ideas conveyed or hinted at in their online communications to stimulate innovative ideas for new product development or service design (e.g., Andreassen & Streukens, 2009; Antons & Breidbach, 2018). If successful, translating consumers' online voice into business actions may potentially establish competitive advantages based on extensive knowledge about customer and market needs and quick reactions to the changing demand matched to internal quality improvement (Herrmann et al., 2000).

Regarding more proactive firm engagement strategies, companies may benefit from actively engaging in the interactive online communication environment. Business interactions with customers on the Internet are evident to the public and thereby the actions taken by firms may be more revelatory than customer–customer interactions. This suggests that firms develop tactics so as to increase visibility and effectiveness of online communication efforts. For example, firms may open up online social platforms such as setting up social media accounts or joining in with third-party-provided programmes to foster online communities and facilitate customers' information exchange. A more direct form of firm–customer interaction is chatting with customers via online communication channels such as social networking sites, online discussion forums, and review systems to disseminate information, provide customer support, and deal with customers' comments and queries. The three-way communication network (Wei et al., 2013) enables firms to voice, and thereby potentially control, eWOM effects by intervening, influencing and mediating the interactions between focal and future customers. Firms and marketers should thoroughly plan social media strategies and guarantee that firms have sufficient exposure to social media. To increase business online visibility, the frequency and timing of posting on social media are critical and expected to have a positive impact on customers' reaction and engagement behaviour. Moreover, communication styles of firm-generated messages have the power to impress customers and, thus, their perception of business trustworthiness. A positive and conversational tone is of paramount importance in generating content. The above strategies can help firms establish a closer relationship with customers and promote the implementation of marketing projects. One point to note is that the extent to which firms should be proactively engaging in OSIs will depend on specific business objectives and managerial needs. However, at the very least, firms should make a move on the Internet by initiating social media activities to pursue competitive advantages.

In addition, providers of online communication platforms can build on the insights developed from this research to design and improve their systems. To encourage more firms and customers to engage in OSIs, these sites can allow firms to respond to customer comments and even support further replies from customers to managerial messages. As such, firms and customers can get into conversations, thereby increasing the likelihood of further visits and usage of the online communication sites. In addition, web designers can develop algorithms to bring such conversations to front pages and highlight favourable messages. Perceptions of the visibility of firms and the quality of online communications can be improved through 'promoting' positive and beneficial firm-generated messages to the public.

Finally, findings from this research imply that the big data approach has potential to derive better insights and enhance accuracy and efficiency in the decision-making process by capitalising on substantial data from market and end consumers. A demanding variant for organisations would be to commit to a data-driven culture. Companies must be aware of the changing environment and the value of big data so as to develop strategic planning for mobilising it. Top management should note that investment in information technology and data talent is the real foundation for enhancing big data competencies. Companies can invest in information technology to implement IT functionality and improve the capability of IT systems to collect, store and process big data. To equip organisations with big data analytics ability, data scientists and business analysts are particularly important. They have technical know-how of advanced analytics and programming skills and can provide data insights for decision makers to find solutions to particular problems. Moreover, enterprising managers can develop strategic plans for social media marketing and put new strategies into action. For example, they may hire social media marketing specialists to manage social media platforms and communities and communicate brands by creating and distributing information. Notably, these practices should be routinised to ensure consistency of management and continuous knowledge development. The above recommended investment will generate payback by creating sustainable benefits in the dynamic business landscape. A closing remark should address the privacy issues related to big data. As discussed in Yadav et al. (2013), trust between firms and customers is critical to customers' willingness to share information. Organisations should act cautiously when collecting and using consumer data, and progress on big data policies would benefit the realisation of big data value.

6.4 Limitations and Future Research

This research reveals many promising avenues for management scholars to explore in the future. As pointed out in the literature review chapter, big data research requires interdisciplinary efforts to unlock its potential value. There are a number of areas in multiple management disciplines, as outlined in Section 2.4.6 of Chapter 2. A compelling research need relates to theoretical attempts to define concepts, advance models, and synthesise diverse views so as to provide a sound theoretical basis for researchers to disentangle the big data phenomenon. Besides, current research interests among different topics are not evenly distributed and this calls for development of research programmes in all management disciplines to measure the business impact of big data. For instance, in this doctoral research, the focus has been given to UGC (online customer reviews) with a discussion on firm engagement strategies to advance marketing-related strategic planning (specifically in relation to customer experience, customer satisfaction and customer engagement). A possible extension is to explore the value of UGC from other management perspectives—for example, how to stimulate sustained innovation through crowdsourcing or how customer behaviour and opinions affect employees' job performance. These directions may lead to fruitful discussions on how to strategise about big data in the rapidly changing context.

Investigations into firm engagement in OSIs in this doctoral research have focused on firm–customer interactions. Yadav and Pavlou (2014) suggest that interactions also exist between firms, but this issue has received limited attention in the current literature. Accordingly, the integrative framework of firm engagement in OSIs proposed in Section 1.4 of Chapter 1 can be extended by taking into consideration competitive interactions between rival firms. For example, rivals' actions may also be a driving force for firm engagement, which may in turn influence competitors' behaviour and the market environment, as implied by a strategic conflict approach (Shapiro, 1989). This could potentially lead to an interesting discussion on the conflict or cooperation among businesses through a game-theoretic lens and how the business benefits from engaging in OSIs change with different equilibrium strategies.

In addition, findings from the three empirical studies can be further validated and generalised with new data and alternative methods. For instance, this doctoral research uses online customer reviews and managerial responses from the hotel sector to analyse effects of firm engagement. This specific research setting can be expanded to other OSI platforms such as Twitter, which has a quicker flow of messages and interactions. It is also well worth testing the effects of firm engagement using data from other service organisations or industrial sectors such as retailing, education and healthcare. It would be interesting to compare the results to appreciate the differences in organisation engagement impacts across different online communication platforms and across different sectors. In addition, specific data characteristics require the application of specific models and methods to analyse and tease out insights from diversified sources and types of data. Strategies driven by big data can be further validated by using traditional analytics approaches such as survey or interview to complement our understanding of this business problem. Overall, the recommended extensions of the thesis are expected to lead to fruitful research on big data-led strategy development and improvement.

REFERENCES

Abadie, A., Drukker, D., Herr, J. L., & Imbens, G. W. (2004). Implementing matching estimators for average treatment effects in Stata. *Stata Journal*, 4, 290–311.

Abrahams, A. S., Fan, W., Wang, G. A., Zhang, Z. J., & Jiao, J. (2015). An integrated text analytic framework for product defect discovery. *Production and Operations Management*, 24(6), 975–990.

Akbaba, A. (2006). Measuring service quality in the hotel industry: A study in a business hotel in Turkey. *International Journal of Hospitality Management*, 25(2), 170–192.

Akter, S., Fosso Wamba, S., Gunasekaran, A., Dubey, R., & Childe, S. J. (2016). How to improve firm performance using big data analytics capability and business strategy alignment?. *International Journal of Production Economics*, 182, 113–131.

Alfaro, C., Cano-Montero, J., Gómez, J., Moguerza, J. M., & Ortega, F. (2016). A multi-stage method for content classification and opinion mining on weblog comments. *Annals of Operations Research*, 236(1), 197–213.

Alharthi, A., Krotov, V., & Bowman, M. (2017). Addressing barriers to big data. *Business Horizons*, 60(3), 285–292.

Aliguliyev, R. (2009a). A new sentence similarity measure and sentence based extractive technique for automatic text summarization. *Expert Systems with Applications*, 36(4), 7764–7772.

Aliguliyev, R. (2009b). Clustering of document collection—A weighting approach. *Expert Systems with Applications*, 36(4), 7904–7916.

Aloysius, J. A., Hoehle, H., & Venkatesh, V. (2016). Exploiting big data for customer and retailer benefits: A study of emerging mobile checkout scenarios. *International Journal of Operations & Production Management,* 36(4), 467–486.

Amankwah-Amoah, J. (2015). Safety or no safety in numbers? Governments, big data and public policy formulation. *Industrial Management and Data System*, 115(9), 1596–1603.

Amankwah-Amoah, J. (2016). Emerging economies, emerging challenges: Mobilising and capturing value from big data. *Technological Forecasting and Social Change*, 10, 167–174.

Amaro, S., Duarte, P., & Henriques, C. (2016). Travelers' use of social media: A clustering approach. *Annals of Tourism Research*, 59, 1–15.

Anderson, E. W., & Sullivan, M. W. (1993). The antecedents and consequences of customer satisfaction for firms. *Marketing Science*, 12(2), 125–143.

Andreassen, T. W., & Streukens, S. (2009). Service innovation and electronic word-of-mouth: Is it worth listening to? *Managing Service Quality: An International Journal*, 19(3), 249–265.

Andrews, M., Luo, X., Fang, Z., & Ghose, A. (2016). Mobile ad effectiveness: Hyper-contextual targeting with crowdedness. *Marketing Science*, 35(2), 218–233.

Antons, D., & Breidbach, C. F. (2018). Big data, big insights? Advancing service innovation and design with machine learning. *Journal of Service Research*, 21(1), 17–39.

Antony, J., Antony, F. J., & Ghosh, S. (2004). Evaluating service quality in a UK hotel chain: A case study. *International Journal of Contemporary Hospitality Management*, 16(6), 380–384.

Archak, N., Ghose, A., & Ipeirotis, P. (2011). Deriving the pricing power of product features by mining consumer reviews. *Management Science*, 57(8), 1485–1509.

Arndt, J. (1967). Word-of-mouth advertising and informal communication. In D. F. Cox (Ed.), Risk taking and information handling in consumer behavior (pp. 188–239). Cambridge: Harvard University Press.

Ashley, C., & Tuten, T. (2015). Creative strategies in social media marketing: An exploratory study of branded social content and consumer engagement. *Psychology & Marketing*, 32(1), 15–27.

Ashton, T., Evangelopoulos, N., & Prybutok, V. (2014). Extending monitoring methods to textual data: A research agenda. *Quality & Quantity*, 48(4), 2277–2294.

Assunção, M., Calheiros, R., Bianchi, S., Netto, M., & Buyya, R. (2015). Big data computing and clouds: Trends and future directions. *Journal of Parallel and Distributed Computing*, 79–80, 3–15.

Ayeh, J. K., Au, N., & Law, R. (2013). "Do we believe in TripAdvisor?" Examining credibility perceptions and online travelers' attitude toward using user-generated content. *Journal of Travel Research*, 52(4), 437–452.

Baak, A., Müller, M., Bharaj, G., Seidel, H. P., & Theobalt, C. (2013). A data-driven approach for real-time full body pose reconstruction from a depth camera. In *Consumer Depth Cameras for Computer Vision* (pp. 71–98). Springer London.

Babić Rosario, A., Sotgiu, F., De Valck, K., & Bijmolt, T. H. (2016). The effect of electronic word of mouth on sales: A meta-analytic review of platform, product, and metric factors. *Journal of Marketing Research*, 53(3), 297–318.

Baek, H., Ahn, J., & Choi, Y. (2012). Helpfulness of online consumer reviews: Readers' objectives and review cues. *International Journal of Electronic Commerce*, 17(2), 99–126.

Balahur, A., Hermida, J., & Montoyo, A. (2012). Detecting implicit expressions of emotion in text: A comparative analysis. *Decision Support Systems*, 53(4), 742–753.

Balaji, M. S., Khong, K. W., & Chong, A. Y. L. (2016). Determinants of negative word-of-mouth communication using social networking sites. *Information & Management*, 53(4), 528–540.

Balakrishnan, R., Qiu, X. Y., & Srinivasan, P. (2010). On the predictive ability of narrative disclosures in annual reports. *European Journal of Operational Research*, 202(3), 789–801.

Banerjee, S., & Chua, A. Y. (2016). In search of patterns among travellers' hotel ratings in TripAdvisor. *Tourism Management*, 53, 125–131.

Bao, Y., & Datta, A. (2014). Simultaneously discovering and quantifying risk types from textual risk disclosures. *Management Science*, 60(6), 1371–1391.

Baralis, E., Cagliero, L., Jabeen, S., Fiori, A., & Shah, S. (2013). Multi-document summarization based on the Yago ontology. *Expert Systems with Applications*, 40(17), 6976–6984.

Barker, R., & Massiglia, P. (2002). *Storage area network essentials*. New York, NY: Wiley.

Barrales-Molina, V., Martínez-López, F. J., & Gázquez-Abad, J. C. (2014). Dynamic marketing capabilities: Toward an integrative framework. *International Journal of Management Reviews*, 16(4), 397–416.

Barroso, L., Clidaras, J., & Hölzle, U. (2013). The datacenter as a computer: An introduction to the design of warehouse-scale machines. *Synthesis Lectures on Computer Architecture*, 8(3), 1–154.

Barua, A., Thomas, S. W., & Hassan, A. E. (2014). What are developers talking about? An analysis of topics and trends in stack overflow. *Empirical Software Engineering*, 19(3), 619–654.

Beath, C., Becerra-Fernandez, I., Ross, J., & Short, J. (2012). Finding value in the information explosion. *MIT Sloan Management Review*, 53(4), 18–20.

Beebe, N., Clark, J., Dietrich, G., Ko, M., & Ko, D. (2011). Post-retrieval search hit clustering to improve information retrieval effectiveness: Two digital forensics case studies. *Decision Support Systems*, 51(4), 732–744.

Berger, C. R., & Calabrese, R. J. (1975). Some explorations in initial interaction and beyond: Toward a developmental theory of interpersonal communication. *Human Communication Research*, 1(2), 99–112.

Berger, J. (2014). Word of mouth and interpersonal communication: A review and directions for future research. *Journal of Consumer Psychology*, 24(4), 586–607.

Berry, L. L., Shankar, V., Parish, J. T., Cadwallader, S., & Dotzel, T. (2006). Creating new markets through service innovation. *MIT Sloan Management Review*, 47(2), 56–63.

Berry, L. L., Zeithaml, V. A., & Parasuraman, A. (1985). Quality counts in services, too. *Business Horizons*, 28(3), 44–52.

Beukeboom, C., Kerkhof, P., & De Vries, M. (2015). Does a virtual like cause actual liking? How following a brand's Facebook updates enhances brand evaluations and purchase intention. *Journal of Interactive Marketing*, 32, 26–36.

Bharadwaj, A., El Sawy, O. A., Pavlou, P. A., & Venkatraman, N. (2013). Digital business strategy: Toward a next generation of insights. *MIS Quarterly*, 37(2), 471–482.

Bharadwaj, N., & Noble, C. H. (2015). Innovation in data-rich environments. *Journal of Product Innovation Management*, 32(3), 476–478.

Bharadwaj, N., Nevin, J. R., & Wallman, J. P. (2012). Explicating hearing the voice of the customer as a manifestation of customer focus and assessing its consequences. *Journal of Product Innovation Management*, 29(6), 1012–1030.

Bhimani, A. (2015). Exploring big data's strategic consequences. *Journal of Information Technology*, 30(1), 66–69.

Bi, Z., & Cochran, D. (2014). Big data analytics with applications. *Journal of Management Analytics*, 1(4), 249–265.

Bies, R., & Moag, R. (1986). Interactional justice: Communication criteria of fairness. In R. J. Lewicki, B. H. Sheppard & M. H. Bazerman (Eds.), *Research on negotiations in organizations* (pp. 43–55). Greenwich: JAI Press.

Bird, B. R., & Smith, E. A. (2005). Signaling theory, strategic interaction, and symbolic capital. *Current Anthropology*, 46(2), 221–248.

Blei, D. M. (2012). Probabilistic topic models. *Communications of the ACM*, 55(4), 77–84.

Blei, D. M., & Lafferty, J. D. (2009). Topic models. In A. N. Srivastava & M. Sahami (Eds.), *Text mining: Classification, clustering, and applications* (pp. 71–94). Boca Raton: CRC Press.

Blei, D. M., Ng, A. Y., & Jordan, M. I. (2003). Latent dirichlet allocation. *Journal of Machine Learning Research*, 3, 993–1022.

Bollier, D., & Firestone, C. M. (2010). *The promise and peril of big data.* Washington, DC: Aspen Institute, Communications and Society Program.

Bolton, R. N. (2011). Comment: Customer engagement: Opportunities and challenges for organizations. *Journal of Service Research*, 14(3), 272–274.

Braganza, A., Brooks, L., Nepelski, D., Ali, M., & Moro, R. (2017). Resource management in big data initiatives: Processes and dynamic capabilities. *Journal of Business Research*, 70, 328–337.

Brown, B., Chui, M., & Manyika, J. (2011). Are you ready for the era of 'big data'. *McKinsey Quarterly*, 4, 24–35.

Brown-Liburd, H., Issa, H., & Lombardi, D. (2015). Behavioral implications of big data's impact on audit judgment and decision making and future research directions. *Accounting Horizons*, 29(2), 451–468.

Bryman, A., & Bell, E. (2015). *Business research methods.* Oxford University Press, USA.

Buechel, E. C., & Berger, J. (2018). Microblogging and the value of undirected communication. *Journal of Consumer Psychology*, 28(1), 40–55.

Bughin, J., Chui, M., & Manyika, J. (2010). Clouds, big data, and smart assets: Ten tech–enabled business trends to watch. *McKinsey Quarterly*, 56(1), 75–86.

Bughin, J., Livingston, J., & Marwaha, S. (2011). Seizing the potential of 'big data'. *McKinsey Quarterly*, 4, 103–109.

Büschken, J., & Allenby, G. M. (2016). Sentence-based text analysis for customer reviews. *Marketing Science*, 35(6), 953–975.

Cachia, R., Compañó, R., & Da Costa, O. (2007). Grasping the potential of online social networks for foresight. *Technological Forecasting and Social Change*, 74(8), 1179–1203.

Cadotte, E. R., & Turgeon, N. (1988). Key factors in guest satisfaction. *Cornell Hotel and Restaurant Administration Quarterly*, 28(4), 44–51.

Caliendo, M., & Kopeinig, S. (2008). Some practical guidance for the implementation of propensity score matching. *Journal of Economic Surveys*, 22(1), 31–72.

Camiciottoli, B. C., Ranfagni, S., & Guercini, S. (2014). Exploring brand associations: An innovative methodological approach. *European Journal of Marketing*, 48(5/6), 1092–1112.

Cantallops, A.S., & Salvi, F. (2014). New consumer behavior: A review of research on eWOM and hotels. *International Journal of Hospitality Management*, 36, 41–51.

Cao, M., Chychyla, R., & Stewart, T. (2015). Big data analytics in financial statement audits. *Accounting Horizons*, 29(2), 423–429.

Cao, Q., Duan, W., & Gan, Q. (2011). Exploring determinants of voting for the "helpfulness" of online user reviews: A text mining approach. *Decision Support Systems*, 50(2), 511–521.

Cascio, C. N., O'Donnell, M. B., Bayer, J., Tinney Jr, F. J., & Falk, E. B. (2015). Neural correlates of susceptibility to group opinions in online word-of-mouth recommendations. *Journal of Marketing Research*, 52(4), 559–575.

Castellanos, M., Gupta, C., Wang, S., Dayal, U., & Durazo, M. (2012). A platform for situational awareness in operational BI. *Decision Support Systems*, 52(4), 869–883.

Cattell, R. (2011). Scalable SQL and NoSQL data stores. *ACM SIGMOD Record*, 39(4), 12–27.

Cavusgil, S. T., Knight, G. A., & Riesenberger, J. R. (2012). *International business: Strategy, management, and the new realitie*s (2nd ed.). Upper Saddle River: Pearson Prentice Hall.

Chae, B. K. (2015). Insights from hashtag# supplychain and Twitter analytics: Considering Twitter and Twitter data for supply chain practice and research. *International Journal of Production Economics*, 165, 247–259.

Chan, H. K., Lacka, E., Yee, R. W., & Lim, M. K. (2017). The role of social media data in operations and production management. *International Journal of Production Research*, 55(17), 5027–5036.

Chan, H. K., Wang, X., Lacka, E., & Zhang, M. (2016). A mixed-method approach to extracting the value of social media data. *Production and Operations Management*, 25(3), 568–583.

Chang, Y. T., Yu, H., & Lu, H. P. (2015). Persuasive messages, popularity cohesion, and message diffusion in social media marketing. *Journal of Business Research*, 68(4), 777–782.

Chatterjee, P., Hoffman, D., & Novak, T. (2003). Modeling the clickstream: Implications for web-based advertising efforts. *Marketing Science*, 22(4), 520–541.

Chau, M., & Chen, H. (2008). A machine learning approach to web page filtering using content and structure analysis. *Decision Support Systems*, 44(2), 482–494.

Chau, M., & Xu, J. (2007). Mining communities and their relationships in blogs: A study of online hate groups. *International Journal of Human–Computer Studies*, 65(1), 57–70.

Chaudhuri, S., Dayal, U., & Narasayya, V. (2011). An overview of business intelligence technology. *Communications of the ACM*, 54(8), 88–98.

Chaxel, A. S., & Han, Y. (2018). Benefiting from disagreement: Counterarguing reduces prechoice bias in information evaluation. *Journal of Consumer Psychology*, 28(1), 115–129.

Chen, H., Chiang, R. H., & Storey, V. C. (2012). Business intelligence and analytics: From big data to big impact. *MIS Quarterly*, 36(4), 1165–1188.

Chen, H., Zheng, Z. E., & Ceran, Y. (2016). De-biasing the reporting bias in social media analytics. *Production and Operations Management*, 25(5), 849–865.

Chen, M., Mao, S., & Liu, Y. (2014). Big data: a survey. *Mobile Networks and Applications*, 19(2), 171–209.

Chen, Y., & Xie, J. (2005). Third-party product review and firm marketing strategy. *Marketing Science*, 24(2), 218–240.

Chen, Z. Y., Fan, Z. P., & Sun, M. (2015). Behavior-aware user response modeling in social media: Learning from diverse heterogeneous data. *European Journal of Operational Research*, 241(2), 422–434.

Cheng, Y. H., & Ho, H. Y. (2015). Social influence's impact on reader perceptions of online reviews. *Journal of Business Research*, 68(4), 883–887.

Cherniack, M., Balakrishnan, H., Balazinska, M., Carney, D., Cetintemel, U., Xing, Y., & Zdonik, S. B. (2003). Scalable Distributed Stream Processing. In *CIDR*, 3 (pp. 257–268).

Cheung, C. M., Xiao, B. S., & Liu, I. L. (2014). Do actions speak louder than voices? The signaling role of social information cues in influencing consumer purchase decisions. *Decision Support Systems*, 65, 50–58.

Chevalier, J. A., & Mayzlin, D. (2006). The effect of word of mouth on sales: Online book reviews. *Journal of Marketing Research*, 43(3), 345–354.

Chevalier, J. A., Dover, Y., & Mayzlin, D. (2017). Channels of impact: User reviews when quality is dynamic and managers respond. Working Paper, No. w23299, National Bureau of Economic Research.

Chodorow, K. (2013). *MongoDB: the definitive guide*. CA: O'Reilly Media, Inc.

Chong, A. Y. L., Li, B., Ngai, E. W., Ch'ng, E., & Lee, F. (2016). Predicting online product sales via online reviews, sentiments, and promotion strategies: A big data architecture and neural network approach. *International Journal of Operations & Production Management*, 36(4), 358–383.

Chou, C. H., Sinha, A. P., & Zhao, H. (2010). A hybrid attribute selection approach for text classification. *Journal of the Association for Information Systems*, 11(9), 491–518.

Choudhary, S., Dincturk, M. E., Mirtaheri, S. M., Moosavi, A., Von Bochmann, G., Jourdan, G. V., & Onut, I. V. (2012). Crawling rich internet applications: the state of the art. In *Proceedings of the 2012 Conference of the Center for Advanced Studies on Collaborative Research* (pp. 146–160). IBM Corp.

Cattell, R. (2011). Scalable SQL and NoSQL data stores. *ACM SIGMOD Record*, 39(4), 12–27.

Christiansen, L. (2011). Personal privacy and Internet marketing: An impossible conflict or a marriage made in heaven? *Business Horizons*, 54(6), 509–514.

Chu, R. K., & Choi, T. (2000). An importance-performance analysis of hotel selection factors in the Hong Kong hotel industry: A comparison of business and leisure travellers. *Tourism Management*, 21(4), 363–377.

Chung, T. S., Wedel, M., & Rust, R. T. (2016). Adaptive personalization using social networks. *Journal of the Academy of Marketing Science*, 44(1), 66–87.

Chung, W., Chen, H., & Nunamaker Jr, J. F. (2005). A visual framework for knowledge discovery on the Web: An empirical study of business intelligence exploration. *Journal of Management Information Systems*, 21(4), 57–84.

Churchill Jr, G. A., & Surprenant, C. (1982). An investigation into the determinants of customer satisfaction. *Journal of Marketing Research*, 19(4), 491–504.

Claussen, J., Kretschmer, T., & Mayrhofer, P. (2013). The effects of rewarding user engagement: The case of Facebook apps. *Information Systems Research*, 24(1), 186–200.

Colace, F., Casaburi, L., De Santo, M., & Greco, L. (2015). Sentiment detection in social networks and in collaborative learning environments. *Computers in Human Behavior*, 51, 1061–1067.

Colace, F., De Santo, M., Greco, L., & Napoletano, P. (2014). Text classification using a few labeled examples. *Computers in Human Behavior*, 30, 689–697.

Colace, F., De Santo, M., Greco, L., Moscato, V., & Picariello, A. (2015). A collaborative user–centered framework for recommending items in online social networks. *Computers in Human Behavior*, 51, 694–704.

Condie, T., Conway, N., Alvaro, P., Hellerstein, J. M., Elmeleegy, K., & Sears, R. (2010). MapReduce online. In *Nsdi*, 10(4) (pp. 20–33).

Connelly, B. L., Certo, S. T., Ireland, R. D., & Reutzel, C. R. (2011). Signaling theory: A review and assessment. *Journal of Management*, 37(1), 39–67.

Cook, L. S., Bowen, D. E., Chase, R. B., Dasu, S., Stewart, D. M., & Tansik, D. A. (2002). Human issues in service design. *Journal of Operations Management*, 20(2), 159–174.

Côrte-Real, N., Oliveira, T., & Ruivo, P. (2017). Assessing business value of big data analytics in European firms. *Journal of Business Research*, 70, 379–390.

Craig, T., & Ludloff, M. (2011). *Privacy and big data.* CA: O'Reilly Media, Inc.

Crawford, C. M. (2008). *New products management.* Boston: McGraw-Hill Education.

Creswell, J. W. (2013). *Research design: Qualitative, quantitative, and mixed methods approaches.* California: Sage publications.

Cronin Jr, J. J., & Taylor, S. A. (1992). Measuring service quality: A reexamination and extension. *Journal of Marketing*, 55–68.

Culotta, A., & Cutler, J. (2016). Mining brand perceptions from twitter social networks. *Marketing Science*, 35(3), 343–362.

Czinkota, M., Ronkainen, I. A., Sutton-Brady, C., & Beal, T. (2011), *International marketing.* Australia: Cegage Learning.

D'Haen, J., Van den Poel, D., & Thorleuchter, D. (2013). Predicting customer profitability during acquisition: Finding the optimal combination of data source and data mining technique. *Expert Systems with Applications*, 40(6), 2007–2012.

Da Silva, N. F., Hruschka, E. R., & Hruschka Jr, E. R. (2014). Tweet sentiment analysis with classifier ensembles. *Decision Support Systems*, 66, 170–179.

Dahl, D. W., Fuchs, C., & Schreier, M. (2015). Why and when consumers prefer products of user-driven firms: A social identification account. *Management Science*, 61(8), 1978–1988.

Das, S. R., & Chen, M. Y. (2007). Yahoo! for Amazon: Sentiment extraction from small talk on the web. *Management Science*, 53(9), 1375–1388.

D'Aveni, R. A., Dagnino, G. B., & Smith, K. G. (2010). The age of temporary advantage. *Strategic Management Journal*, 31(13), 1371–1385.

Davenport, T. H. (2006). Competing on analytics. *Harvard Business Review*, 84(1), 98–107.

Davenport, T. H., Barth, P., & Bean, R. (2012). How big data is different. *MIT Sloan Management Review*, 54(1), 43–46.

Davidow, M. (2003). Organizational responses to customer complaints: What works and what doesn't. *Journal of Service Research*, 5(3), 225–250.

De Matos, C. A., & Rossi, C. A. V. (2008). Word-of-mouth communications in marketing: A meta-analytic review of the antecedents and moderators. *Journal of the Academy of Marketing Science*, 36(4), 578–596.

De Vries, L., Gensler, S., & Leeflang, P. S. (2012). Popularity of brand posts on brand fan pages: An investigation of the effects of social media marketing. *Journal of Interactive Marketing*, 26(2), 83–91.

De Vries, L., Gensler, S., & Leeflang, P. S. (2017). Effects of traditional advertising and social messages on brand-building metrics and customer acquisition. *Journal of Marketing*, 81(5), 1–15.

Dean, J., & Ghemawat, S. (2008). MapReduce: Simplified data processing on large clusters. *Communications of the ACM*, 51(1), 107–113.

Dean, J., & Ghemawat, S. (2010). MapReduce: A flexible data processing tool. *Communications of the ACM*, 53(1), 72–77.

DeCandia, G., Hastorun, D., Jampani, M., Kakulapati, G., Lakshman, A., Pilchin, A., ... & Vogels, W. (2007). Dynamo: Amazon's highly available key-value store. *ACM SIGOPS Operating Systems Review*, 41(6), 205–220.

Deephouse, D. L. (2000). Media reputation as a strategic resource: An integration of mass communication and resource-based theories. *Journal of Management*, 26(6), 1091–1112.

Deerwester, S., Dumais, S. T., Furnas, G. W., Landauer, T. K., & Harshman, R. (1990). Indexing by latent semantic analysis. *Journal of the American Society for Information Science*, 41(6), 391–407.

Dehejia, R. H., & Wahba, S. (2002). Propensity score-matching methods for nonexperimental causal studies. *The Review of Economics and Statistics*, 84(1), 151–161.

Delen, D., & Demirkan, H. (2013). Data, information and analytics as services. *Decision Support Systems*, 55(1), 359–363.

Dellarocas, C. (2006). Strategic manipulation of internet opinion forums: Implications for consumers and firms. *Management Science*, 52(10), 1577–1593.

Dellarocas, C., Zhang, X., & Awad, N. (2007). Exploring the value of online product reviews in forecasting sales: The case of motion pictures. *Journal of Interactive Marketing*, 21(4), 23–45.

Dijkmans, C., Kerkhof, P., & Beukeboom, C. J. (2015). A stage to engage: Social media use and corporate reputation. *Tourism Management*, 47, 58–67.

Ding, A. W., Li, S., & Chatterjee, P. (2015). Learning user real-time intent for optimal dynamic web page transformation. *Information Systems Research*, 26(2), 339–359.

Ding, D., Metze, F., Rawat, S., Schulam, P. F., Burger, S., Younessian, E., ... & Hauptmann, A. (2012). Beyond audio and video retrieval: Towards multimedia summarization.

In *Proceedings of the 2nd ACM International Conference on Multimedia Retrieval* (Article No. 2). ACM.

Dong, B., & Sivakumar, K. (2017). Customer participation in services: Domain, scope, and boundaries. *Journal of the Academy of Marketing Science*, 45(6), 944–965.

Douglas, M. (2013). Big data raises big questions. *Government Technology,* 26(4), 12–16.

Duan, J. Y., Zhang, M., Wang, J. Z., & Xu, Y. S. (2011). A hybrid framework to extract bilingual multiword expression from free text. *Expert Systems with Applications*, 38(1), 314–320.

Dutta, D., & Bose, I. (2015). Managing a big data project: The case of Ramco Cements Limited. *International Journal of Production Economics*, 165, 293–306.

Economist Intelligence Unit. (2011). *Big data: Harnessing a game-changing asset.* London: EIU.

Eisingerich, A. B., Chun, H. H., Liu, Y., Jia, H. M., & Bell, S. J. (2015). Why recommend a brand face-to-face but not on Facebook? How word-of-mouth on online social sites differs from traditional word-of-mouth. *Journal of Consumer Psychology*, 25(1), 120–128.

Elgendy, N., & Elragal, A. (2014). Big data analytics: A literature review paper. In *Industrial Conference on Data Mining* (pp. 214–227). Springer International Publishing.

Enz, C. A., & Grover, R. A. (1992). The importance of top management visibility for service-based professionals. *Journal of Managerial Issues*, 414–423.

Erevelles, S., Fukawa, N., & Swayne, L. (2016). Big data consumer analytics and the transformation of marketing. *Journal of Business Research*, 69(2), 897–904.

Ernst, H., Hoyer, W. D., Krafft, M., & Soll, J. H. (2010). Consumer idea generation. Working paper. WHU, Vallendar.

Evans, D. (2010). *Social media marketing: The next generation of business engagement.* Indiana: John Wiley & Sons.

Fan, W., Gordon, M. D., & Pathak, P. (2006). An integrated two-stage model for intelligent information routing. *Decision Support Systems*, 42(1), 362–374.

Fan, W., Wallace, L., Rich, S., & Zhang, Z. (2006). Tapping the power of text mining. *Communications of the ACM*, 49(9), 76–82.

Fang, X., Hu, P. J. H., Li, Z., & Tsai, W. (2013). Predicting adoption probabilities in social networks. *Information Systems Research*, 24(1), 128–145.

Feldman, R., & Sanger, J. (2007). *The text mining handbook: Advanced approaches in analyzing unstructured data.* New York, NY: Cambridge University Press.

Felix, R., Rauschnabel, P. A., & Hinsch, C. (2017). Elements of strategic social media marketing: A holistic framework. *Journal of Business Research*, 70, 118–126.

Feng, H., Tian, J., Wang, H. J., & Li, M. (2015). Personalized recommendations based on time-weighted overlapping community detection. *Information & Management*, 52(7), 789–800.

Flint, D. J., Blocker, C. P., & Boutin Jr, P. J. (2011). Customer value anticipation, customer satisfaction and loyalty: An empirical examination. *Industrial Marketing Management*, 40(2), 219–230.

Fong, N. M., Fang, Z., & Luo, X. (2015). Geo-conquesting: Competitive locational targeting of mobile promotions. *Journal of Marketing Research*, 52(5), 726–735.

Fosso Wamba, S., Akter, S., Edwards, A., Chopin, G., & Gnanzou, D. (2015). How 'big data' can make big impact: Findings from a systematic review and a longitudinal case study. *International Journal of Production Economics*, 165, 234–246.

Fosso Wamba, S., Gunasekaran, A., Akter, S., Ren, S. J. F., Dubey, R., & Childe, S. J. (2017). Big data analytics and firm performance: Effects of dynamic capabilities. *Journal of Business Research*, 70, 356–365.

France, S. L., & Ghose, S. (2016). An analysis and visualization methodology for identifying and testing market structure. *Marketing Science*, 35(1), 182–197.

Franks, B. (2012). *Taming the big data tidal wave: Finding opportunities in huge data streams with advanced analytics.* New York, NY: John Wiley.

Friedkin, N. E. (1998). *A structural theory of social influence.* NY: Cambridge University Press.

Fuchs, C., & Otto, A. (2015). Value of IT in supply chain planning. *Journal of Enterprise Information Management*, 28(1), 77–92.

Fuchs, C., & Schreier, M. (2011). Customer empowerment in new product development. *Journal of Product Innovation Management*, 28(1), 17–32.

Fuchs, C., Prandelli, E., & Schreier, M. (2010). The psychological effects of empowerment strategies on consumers' product demand. *Journal of Marketing*, 74(1), 65–79.

Fuchs, C., Prandelli, E., Schreier, M., & Dahl, D. W. (2013). All that is users might not be gold: How labeling products as user designed backfires in the context of luxury fashion brands. *Journal of Marketing*, 77(5), 75–91.

Gandomi, A., & Haider, M. (2015). Beyond the hype: Big data concepts, methods, and analytics. *International Journal of Information Management*, 35(2), 137–144.

Gao, G. G., Greenwood, B. N., Agarwal, R., & Jeffrey, S. (2015). Vocal minority and silent majority: How do online ratings reflect population perceptions of quality?. *MIS Quarterly*, 39(3), 565–589.

Gao, K., Xu, H., & Wang, J. (2015). A rule-based approach to emotion cause detection for Chinese micro-blogs. *Expert Systems with Applications*, 42(9), 4517–4528.

García-Cumbreras, M. Á., Montejo-Ráez, A., & Díaz-Galiano, M. C. (2013). Pessimists and optimists: Improving collaborative filtering through sentiment analysis. *Expert Systems with Applications*, 40(17), 6758–6765.

Garg, R., Smith, M., & Telang, R. (2011). Measuring information diffusion in an online community. *Journal of Management Information Systems*, 28(2), 11–38.

Garg, Y., & Chatterjee, N. (2014). Sentiment analysis of twitter feeds. In *International Conference on Big Data Analytics* (pp. 33–52). Springer International Publishing.

Gelman, A., & Hill, J. (2007). *Data analysis using regression and multilevel/hierarchical models.* NY: Cambridge University Press.

Gensler, S., Völckner, F., Liu-Thompkins, Y., & Wiertz, C. (2013). Managing brands in the social media environment. *Journal of Interactive Marketing*, 27(4), 242–256.

Gerhardt, B., Griffin, K., & Klemann, R. (2012). *Unlocking value in the fragmented world of big data analytics: How information infomediaries will create a new data ecosystem.* Cisco Internet Business Solutions Group.

Ghani, N., Dixit, S., & Wang, T. S. (2000). On IP-over-WDM integration. *IEEE Communications Magazine*, 38(3), 72–84.

Ghemawat, S., Gobioff, H., & Leung, S. (2003). The Google file system. *ACM SIGOPS Operating Systems Review*, 37(5), 29–43.

Ghose, A., & Han, S. (2011). An empirical analysis of user content generation and usage behavior on the mobile internet. *Management Science*, 57(9), 1671–1691.

Ghose, A., & Han, S. (2014). Estimating demand for mobile applications in the new economy. *Management Science*, 60(6), 1470–1488.

Ghose, A., & Todri, V. (2016). Towards a digital attribution model: Measuring the impact of display advertising on online consumer behavior. *MIS Quarterly*, 40(4), 889–910.

Ghose, A., Goldfarb, A., & Han, S. (2013). How is the mobile internet different? Search costs and local activities. *Information Systems Research*, 24(3), 613–631.

Ghose, A., Ipeirotis, P. G., & Li, B. (2012). Designing ranking systems for hotels on travel search engines by mining user-generated and crowdsourced content. *Marketing Science*, 31(3), 493–520.

Gibson, G., & Van Meter, R. (2000). Network attached storage architecture. *Communications of the ACM*, 43(11), 37–45.

Glancy, F., & Yadav, S. (2011). A computational model for financial reporting fraud detection. *Decision Support Systems*, 50(3), 595–601.

Godes, D., & Mayzlin, D. (2009). Firm-created word-of-mouth communication: Evidence from a field test. *Marketing Science*, 28(4), 721–739.

Godes, D., Mayzlin, D., Chen, Y., Das, S., Dellarocas, C., Pfeiffer, B., ... & Verlegh, P. (2005). The firm's management of social interactions. *Marketing Letters*, 16(3–4), 415–428.

Godey, B., Manthiou, A., Pederzoli, D., Rokka, J., Aiello, G., Donvito, R., & Singh, R. (2016). Social media marketing efforts of luxury brands: Influence on brand equity and consumer behavior. *Journal of Business Research*, 69(12), 5833–5841.

Goes, P. B., Lin, M., & Au Yeung, C. M. (2014). 'Popularity effect' in user-generated content: Evidence from online product reviews. *Information Systems Research*, 25(2), 222–238.

Goh, K. Y., Heng, C. S., & Lin, Z. (2013). Social media brand community and consumer behavior: Quantifying the relative impact of user-and marketer-generated content. *Information Systems Research*, 24(1), 88–107.

Goldfarb, A., & Tucker, C. (2011). Privacy regulation and online advertising. *Management Science,* 57(1), 57–71.

Gong, S., Zhang, J., Zhao, P., & Jiang, X. (2017). Tweeting as a marketing tool-field experiment in the TV industry. *Journal of Marketing Research*, 54(6), 833–850.

Gopaldas, A. (2014). Marketplace sentiments. *Journal of Consumer Research*, 41(4), 995–1014.

Gopinath, S., Chintagunta, P. K., & Venkataraman, S. (2013). Blogs, advertising, and local-market movie box office performance. *Management Science*, 59(12), 2635–2654.

Graham, S., Weingart S., & Milligan I. (2002). Getting started with topic modelling and MALLET, available at http://programminghistorian.org/lessons/topic-modeling-and-mallet (accessed 22 November 2016).

Grant, S., & Cordy, J. R. (2010). Estimating the optimal number of latent concepts in source code analysis. In *Source Code Analysis and Manipulation (SCAM) 10th IEEE Working Conference* (pp. 65–74), IEEE.

Gravano, L., Ipeirotis, P. G., Koudas, N., & Srivastava, D. (2003). Text joins in an RDBMS for web data integration. In *Proceedings of the 12th International Conference on World Wide Web* (pp. 90–101). ACM.

Gretry, A., Horváth, C., Belei, N., & van Riel, A. C. (2017). "Don't pretend to be my friend!" When an informal brand communication style backfires on social media. *Journal of Business Research*, 74, 77–89.

Grewal, D., Bart, Y., Spann, M., & Zubcsek, P. P. (2016). Mobile advertising: A framework and research agenda. *Journal of Interactive Marketing*, 34, 3–14.

Griffin, A., & Hauser, J. R. (1993). The voice of the customer. *Marketing Science*, 12(1), 1–27.

Griffiths, T. L., & Steyvers, M. (2004). Finding scientific topics. *Proceedings of the National Academy of Sciences*, 101(suppl 1), 5228–5235.

Grimmelikhuijsen, S. G., & Meijer, A. J. (2015). Does Twitter increase perceived police legitimacy?. *Public Administration Review*, 75(4), 598–607.

Grolinger, K., Hayes, M., Higashino, W. A., L'Heureux, A., Allison, D. S., & Capretz, M. A. (2014). Challenges for mapreduce in big data. In *2014 IEEE World Congress on Services* (pp. 182–189). IEEE.

Gruner, K. E., & Homburg, C. (2000). Does customer interaction enhance new product success?. *Journal of Business Research*, 49(1), 1–14.

Gu, B., & Ye, Q. (2014). First step in social media: Measuring the influence of online management responses on customer satisfaction. *Production and Operations Management*, 23(4), 570–582.

Guesalaga, R. (2016). The use of social media in sales: Individual and organizational antecedents, and the role of customer engagement in social media. *Industrial Marketing Management*, 54, 71–79.

Gunasekaran, A., Papadopoulos, T., Dubey, R., Fosso Wamba, S., Childe, S.J., Hazen, B., & Akter, S. (2017). Big data and predictive analytics for supply chain and organisational performance. *Journal of Business Research*, 70, 308–317.

Guo, Y., Barnes, S. J., & Jia, Q. (2017). Mining meaning from online ratings and reviews: Tourist satisfaction analysis using latent dirichlet allocation. *Tourism Management*, 59, 467–483.

Guo, Z. X., Wong, W. K., & Guo, C. (2014). A cloud-based intelligent decision-making system for order tracking and allocation in apparel manufacturing. *International Journal of Production Research*, 52(4), 1100–1115.

Gupta, M., & George, J. F. (2016). Toward the development of a big data analytics capability. *Information & Management*, 53(8), 1049–1064.

Hahn, G., & Packowski, J. (2015). A perspective on applications of in-memory analytics in supply chain management. *Decision Support Systems*, 76, 45–52.

Hamilton, M., Kaltcheva, V. D., & Rohm, A. J. (2016). Social media and value creation: The role of interaction satisfaction and interaction immersion. *Journal of Interactive Marketing*, 36, 121–133.

Han, J., Kamber, M., & Pei, J. (2011). *Data mining: concepts and techniques: Concepts and techniques* (3rd ed.). MA: Elsevier.

Han, S. P., Park, S., & Oh, W. (2016). Mobile app analytics: A multiple discrete-continuous choice framework. *MIS Quarterly*, 40(4), 983–1008.

Harmeling, C. M., Moffett, J. W., Arnold, M. J., & Carlson, B. D. (2017). Toward a theory of customer engagement marketing. *Journal of the Academy of Marketing Science*, 45(3), 312–335.

Harrigan, P., Evers, U., Miles, M., and Daly, T. (2017). Customer engagement with tourism social media brands. *Tourism Management*, 59, 597–609.

Hashimi, H., Hafez, A., & Mathkour, H. (2015). Selection criteria for text mining approaches. *Computers in Human Behavior*, 51, 729–733.

Hazen, B. T., Boone, C. A., Ezell, J. D., & Jones-Farmer, L. A. (2014). Data quality for data science, predictive analytics, and big data in supply chain management: An introduction to the problem and suggestions for research and applications. *International Journal of Production Economics*, 154, 72–80.

Hazen, B.T., Skipper, J.B., Boone, C.A., & Hill, R.R. (2016). Back in business: Operations research in support of big data analytics for operations and supply chain management. *Annals of Operations Research*, 1–11.

He, W. (2013). Examining students' online interaction in a live video streaming environment using data mining and text mining. *Computers in Human Behavior*, 29(1), 90–102.

Heckmann, J., Ichimura, H., & Todd, P. (1997). Matching as an econometric evaluation estimator. *Review of Economic Studies*, 65(2), 261–294.

Hennig-Thurau, T., Gwinner, K. P., Walsh, G., & Gremler, D. D. (2004). Electronic word-of-mouth via consumer-opinion platforms: What motivates consumers to articulate themselves on the Internet?. *Journal of Interactive Marketing*, 18(1), 38–52.

Hennig-Thurau, T., Malthouse, E. C., Friege, C., Gensler, S., Lobschat, L., Rangaswamy, A., & Skiera, B. (2010). The impact of new media on customer relationships. *Journal of Service Research*, 13(3), 311–330.

Hennig-Thurau, T., Wiertz, C., & Feldhaus, F. (2015). Does Twitter matter? The impact of microblogging word of mouth on consumers' adoption of new movies. *Journal of the Academy of Marketing Science*, 43(3), 375–394.

Herrmann, A., Huber, F., & Braunstein, C. (2000). Market-driven product and service design: Bridging the gap between customer needs, quality management, and customer satisfaction. *International Journal of Production Economics*, 66(1), 77–96.

Higgins, E. T., & Scholer, A. A. (2009). Engaging the consumer: The science and art of the value creation process. *Journal of Consumer Psychology*, 19(2), 100–114.

Hildebrand, C., Häubl, G., Herrmann, A., & Landwehr, J. (2013). When social media can be bad for you: Community feedback stifles consumer creativity and reduces satisfaction with self-designed products. *Information Systems Research*, 24(1), 14–29.

Hitt, M. A., Ireland, R. D., & Hoskisson, R. E. (2011). *Strategic management: Competitiveness and globalization* (9th ed.). Mason, OH: Cengage Learning.

Hitt, M. A., Ireland, R. D., & Hoskisson, R.E. (2014). *Strategic management: Competitiveness and globalization* (11th ed.). Mason, OH: Cengage Learning.

Ho, S. Y., Bodoff, D., & Tam, K. Y. (2011). Timing of adaptive web personalization and its effects on online consumer behavior. *Information Systems Research*, 22(3), 660–679.

Hofmann, M., & Klinkenberg, R. (2013). *RapidMiner: Data mining use cases and business analytics applications*. London: CRC Press.

Hofmann, T. (1999). Probabilistic latent semantic indexing. In *Proceedings of the 22nd Annual International ACM SIGIR Conference on Research and Development in Information Retrieval* (pp. 50–57), ACM.

Hogan, J. E., Lemon, K. N., & Rust, R. T. (2002). Customer equity management: Charting new directions for the future of marketing. *Journal of Service Research*, 5(1), 4–12.

Hollebeek, L. D., Glynn, M. S., & Brodie, R. J. (2014). Consumer brand engagement in social media: Conceptualization, scale development and validation. *Journal of Interactive Marketing*, 28(2), 149–165.

Homburg, C., & Fürst, A. (2007). See no evil, hear no evil, speak no evil: A study of defensive organizational behavior towards customer complaints. *Journal of the Academy of Marketing Science*, 35(4), 523–536.

Homburg, C., Ehm, L., & Artz, M. (2015). Measuring and managing consumer sentiment in an online community environment. *Journal of Marketing Research*, 52(5), 629–641.

Hoyer, W. D., Chandy, R., Dorotic, M., Krafft, M., & Singh, S. S. (2010). Consumer cocreation in new product development. *Journal of Service Research*, 13(3), 283–296.

Hu, H., Wen, Y., Chua, T. S., & Li, X. (2014). Toward scalable systems for big data analytics: A technology tutorial. *IEEE Access,* 2, 652–687.

Hu, N., Bose, I., Koh, N., & Liu, L. (2012). Manipulation of online reviews: An analysis of ratings, readability, and sentiments. *Decision Support Systems*, 52(3), 674–684.

Huang, T., & Van Mieghem, J. A. (2014). Clickstream data and inventory management: Model and empirical analysis. *Production and Operations Management*, 23(3), 333–347.

Huberty, M. (2015). Can we vote with our tweet? On the perennial difficulty of election forecasting with social media. *International Journal of Forecasting*, 31(3), 992–1007.

Hunt, S. D., & Morgan, R. M. (1995). The comparative advantage theory of competition. *Journal of Marketing*, 59(2), 1–15.

Hussain, T., Asghar, S., & Masood, N. (2010). Web usage mining: A survey on preprocessing of web log file. In *2010 International Conference on Information and Emerging Technologies* (pp. 1–6). IEEE.

Hyung, Z., Lee, K., & Lee, K. (2014). Music recommendation using text analysis on song requests to radio stations. *Expert Systems with Applications*, 41(5), 2608–2618.

Ibrahim, N. F., Wang, X., & Bourne, H. (2017). Exploring the effect of user engagement in online brand communities: Evidence from Twitter. *Computers in Human Behavior*, 72, 321–338.

Ippolito, P. M. (1990). Bonding and nonbonding signals of product quality. *Journal of Business*, 63(1), 41–60.

Ireland, R.D., Hoskisson, R.E., & Hitt, M.A. (2012). *Understanding business strategy* (3rd ed.). Cengage Learning.

Iyengar, R., Van den Bulte, C., & Valente, T.W. (2011). Opinion leadership and social contagion in new product diffusion. *Marketing Science*, 30(2), 195–212.

Iyer, G., & Katona, Z. (2016). Competing for attention in social communication markets. *Management Science*, 62(8), 2304–2320.

Järvinen, J., & Karjaluoto, H. (2015). The use of Web analytics for digital marketing performance measurement. *Industrial Marketing Management*, 50, 117–127.

Jeffery, S. R., Alonso, G., Franklin, M. J., Hong, W., & Widom, J. (2006). A pipelined framework for online cleaning of sensor data streams. In *Proceedings of the 22nd International Conference on Data Engineering* (pp. 140–140). IEEE.

Jin, J., Liu, Y., Ji, P., & Liu, H. (2016). Understanding big consumer opinion data for market-driven product design. *International Journal of Production Research*, 54(10), 3019–3041.

Jin, X., Wah, B. W., Cheng, X., & Wang, Y. (2015). Significance and challenges of big data research. *Big Data Research*, 2(2), 59–64.

Johnson, S. L., Safadi, H., & Faraj, S. (2015). The emergence of online community leadership. *Information Systems Research*, 26(1), 165–187.

Jonsson, P., & Eklundh, L. (2002). Seasonality extraction by function fitting to time-series of satellite sensor data. *IEEE Transactions on Geoscience and Remote Sensing*, 40(8), 1824–1832.

Jordan, H., & Alaghband, G. (2002). *Fundamentals of parallel processing.* Upper Saddle River, NJ: Prentice Hall/Pearson Education.

Kache, F., & Seuring, S. (2017). Challenges and opportunities of digital information at the intersection of Big Data Analytics and supply chain management. *International Journal of Operations & Production Management*, 37(1), 10–36.

Kandampully, J., & Suhartanto, D. (2000). Customer loyalty in the hotel industry: The role of customer satisfaction and image. *International Journal of Contemporary Hospitality Management*, 12(6), 346–351.

Kang, D., & Park, Y. (2014). Review-based measurement of customer satisfaction in mobile service: Sentiment analysis and VIKOR approach. *Expert Systems with Applications*, 41(4), 1041–1050.

Kannan, P. K., & Li, H. A. (2017). Digital marketing: A framework, review and research agenda. *International Journal of Research in Marketing*, 34(1), 22–45.

Kao, A., & Poteet, S. (2007). *Natural language processing and text mining.* New York: Springer.

Kaplan, A. M., & Haenlein, M. (2010). Users of the world, unite! The challenges and opportunities of social media. *Business Horizons*, 53(1), 59–68.

Kaplan, E., & Hegarty, C. (2005). *Understanding GPS: Principles and Applications.* MA: Artech house.

Kardes, F. R. (1993). Consumer inference: Determinants, consequences, and implications for advertising. In A. A. Mitchell (Eds.), *Advertising exposure, memory and choice* (pp. 163–191). Hillsdale, NJ: Lawrence Erlbaum Associates.

Kardes, F. R., Posavac, S. S., & Cronley, M. L. (2004). Consumer inference: A review of processes, bases, and judgment contexts. *Journal of Consumer Psychology*, 14(3), 230–256.

Kardes, F. R., Posavac, S. S., Cronley, M. L., & Herr, P. M. (2008). Consumer inference. In C. P. Haugtvedt, P. M. Herr & F. R. Kardes (Eds.), *Handbook of consumer psychology* (pp. 165–192). NY: Lawrence Erlbaum Associates, Psychology Press.

Katal, A., Wazid, M., & Goudar, R. H. (2013). Big data: Issues, challenges, tools and good practices In *Contemporary Computing (IC3) 2013 Sixth International Conference* (pp. 404–409), IEEE.

Khan, F., Bashir, S., & Qamar, U. (2014). TOM: Twitter opinion mining framework using hybrid classification scheme. *Decision Support Systems*, 57, 245–257.

Khan, N., Yaqoob, I., Hashem, I. A. T., Inayat, Z., Mahmoud Ali, W. K., Alam, M., ... & Gani, A. (2014). Big data: Survey, technologies, opportunities, and challenges. *The Scientific World Journal*, 2014, 1–18.

Kim, A. J., & Ko, E. (2012). Do social media marketing activities enhance customer equity? An empirical study of luxury fashion brand. *Journal of Business Research*, 65(10), 1480–1486.

Kim, W. G., Lim, H., & Brymer, R. A. (2015). The effectiveness of managing social media on hotel performance. *International Journal of Hospitality Management*, 44, 165–171.

King, R. A., Racherla, P., & Bush, V. D. (2014). What we know and don't know about online word-of-mouth: A review and synthesis of the literature. *Journal of Interactive Marketing*, 28(3), 167–183.

Kiron, D., & Bean, R. (2013). Organizational alignment is key to big data success. *MIT Sloan Management Review*, 54(3), 1–6.

Kitchin, R. (2014). Big Data, new epistemologies and paradigm shifts. *Big Data & Society*, April–June (1), 1–12.

Kluemper, D. H., & Rosen, P. A. (2009). Future employment selection methods: Evaluating social networking web sites. *Journal of Managerial Psychology*, 24(6), 567–580.

Knutson, B., Stevens, P., Patton, M., & Thompson, C. (1993). Consumers' expectations for service quality in economy, mid-price and luxury hotels. *Journal of Hospitality & Leisure Marketing*, 1(2), 27–43.

Költringer, C., & Dickinger, A. (2015). Analyzing destination branding and image from online sources: A web content mining approach. *Journal of Business Research*, 68(9), 1836–1843.

Kothari, C. R. (2004). *Research methodology: Methods and techniques* (2nd ed.). New Delhi: New Age International.

Kou, G., & Lou, C. (2012). Multiple factor hierarchical clustering algorithm for large scale web page and search engine clickstream data. *Annals of Operations Research*, 197(1), 123–134.

Kozinets, R. V. (2002). The field behind the screen: Using netnography for marketing research in online communities. *Journal of Marketing Research*, 39(1), 61–72.

Kozinets, R. V., De Valck, K., Wojnicki, A. C., & Wilner, S. J. (2010). Networked narratives: Understanding word-of-mouth marketing in online communities. *Journal of Marketing*, 74(2), 71–89.

Kristensson, P., Gustafsson, A., & Archer, T. (2004). Harnessing the creative potential among users. *Journal of Product Innovation Management*, 21(1), 4–14.

Kumar, A., Bezawada, R., Rishika, R., Janakiraman, R., & Kannan, P. K. (2016). From social to sale: The effects of firm-generated content in social media on customer behavior. *Journal of Marketing*, 80(1), 7–25.

Kumar, A., Shankar, R., Choudhary, A., & Thakur, L. S. (2016). A big data MapReduce framework for fault diagnosis in cloud-based manufacturing. *International Journal of Production Research*, 54(23), 7060–7073.

Kumar, V., & Pansari, A. (2016). Competitive advantage through engagement. *Journal of Marketing Research*, 53(4), 497–514.

Kumar, V., Bhaskaran, V., Mirchandani, R., & Shah, M. (2013). Practice prize winner—creating a measurable social media marketing strategy: Increasing the value and ROI of intangibles and tangibles for hokey pokey. *Marketing Science*, 32(2), 194–212.

Kumar, V., Choi, J. B., & Greene, M. (2017). Synergistic effects of social media and traditional marketing on brand sales: Capturing the time-varying effects. *Journal of the Academy of Marketing Science*, 45(2), 268–288.

Kurt, D., Inman, J. J., & Argo, J. J. (2011). The influence of friends on consumer spending: The role of agency-communion orientation and self-monitoring. *Journal of Marketing Research*, 48(4), 741–754.

Kwon, O., Lee, N., & Shin, B. (2014). Data quality management, data usage experience and acquisition intention of big data analytics. *International Journal of Information Management*, 34(3), 387–394.

Ladhari, R. (2009). A review of twenty years of SERVQUAL research. *International Journal of Quality and Service Sciences*, 1(2), 172–198.

Lam, S. K., Sleep, S., Hennig-Thurau, T., Sridhar, S., & Saboo, A. R. (2017). Leveraging frontline employees' small data and firm-level big data in frontline management an absorptive capacity perspective. *Journal of Service Research*, 20(1), 12–28.

Lamberton, C., & Stephen, A. T. (2016). A thematic exploration of digital, social media, and mobile marketing: Research evolution from 2000 to 2015 and an agenda for future inquiry. *Journal of Marketing*, 80(6), 146–172.

Laney, D. (2001). 3-D data management: controlling data volume, velocity and variety, *META Group Research Note*, 6, available at http://blogs.gartner.com/doug-laney/files/2012/01/ad949-3D-Data-Management-Controlling-Data-Volume-Velocity-and-Variety.pdf (accessed 11 March 2015).

Lappas, T., Sabnis, G., & Valkanas, G. (2016). The impact of fake reviews on online visibility: A vulnerability assessment of the hotel industry. *Information Systems Research*, 27(4), 940–961.

Lau, R. Y., Liao, S. S., Wong, K. F., & Chiu, D. K. (2012). Web 2.0 environmental scanning and adaptive decision support for business mergers and acquisitions. *MIS Quarterly*, 36(4), 1239–1268.

Laurila, J. K., Gatica-Perez, D., Aad, I., Bornet, O., Do, T. M. T., Dousse, O., ... & Miettinen, M. (2012). The mobile data challenge: Big data for mobile computing research. In *Pervasive Computing* (No. EPFL-CONF-192489).

LaValle, S., Lesser, E., Shockley, R., Hopkins, M. S., & Kruschwitz, N. (2011). Big data, analytics and the path from insights to value. *MIT Sloan Management Review*, 52(2), 20–32.

Lavertu, S. (2015). We all need help: "Big data" and the mismeasure of public administration, *Public Administration Review*, 76(6), 864–872.

Lee, C. C., & Hu, C. (2005). Analyzing hotel customers' e-complaints from an internet complaint forum. *Journal of Travel & Tourism Marketing*, 17(2), 167–181.

Lee, C. H., & Cranage, D. A. (2014). Toward understanding consumer processing of negative online word-of-mouth communication: The roles of opinion consensus and organizational response strategies. *Journal of Hospitality & Tourism Research*, 38(3), 330–360.

Lee, C., & Wang, S. (2012). An information fusion approach to integrate image annotation and text mining methods for geographic knowledge discovery. *Expert Systems with Applications*, 39(10), 8954–8967.

Lee, C., Yang, H., & Wang, S. (2011). An image annotation approach using location references to enhance geographic knowledge discovery. *Expert Systems with Applications*, 38(11), 13792–13802.

Lee, I. (2017). Big data: Dimensions, evolution, impacts, and challenges. *Business Horizons*, 60, 293–303

Lee, W. (2007). Deploying personalized mobile services in an agent-based environment. *Expert Systems with Applications*, 32(4), 1194–1207.

Lee, T., & BradLow, E. (2011). Automated marketing research using online customer reviews. *Journal of Marketing Research*, 48(5), 881–894.

Lee, Y. J., Hosanagar, K., & Tan, Y. (2015). Do I follow my friends or the crowd? Information cascades in online movie ratings. *Management Science*, 61(9), 2241–2258.

Lee, Y. J., Xie, K., Besharat, A., & Tan, Y. (2017). Management responses to online WOM: H elpful or detrimental? Retrieved from https://papers.ssrn.com/sol3/papers.cfm?abstract_i d=2919553

Lee, Y. L., & Song, S. (2010). An empirical investigation of electronic word-of-mouth: Informational motive and corporate response strategy. *Computers in Human Behavior*, 26(5), 1073–1080.

Leeflang, P., Verhoef, P., Dahlström, P., & Freundt, T. (2014). Challenges and solutions for marketing in a digital era. *European Management Journal*, 32(1), 1–12.

Lenzerini, M. (2002). Data integration: A theoretical perspective. In *Proceedings of the 21st ACM SIGMOD-SIGACT-SIGART Symposium on Principles of Database Systems* (pp. 233–246). ACM.

Lew, M. S., Sebe, N., Djeraba, C., & Jain, R. (2006). Content-based multimedia information retrieval: State of the art and challenges. *ACM Transactions on Multimedia Computing, Communications, and Applications*, 2(1), 1–19.

Li, D., & Wang, X. (2017). Dynamic supply chain decisions based on networked sensor data: An application in the chilled food retail chain. *International Journal of Production Research*, 55(17), 5127–5141.

Li, H., Ye, Q., & Law, R. (2013). Determinants of customer satisfaction in the hotel industry: An application of online review analysis. *Asia Pacific Journal of Tourism Research*, 18(7), 784–802.

Li, J., Wang, H., & Bai, X. (2015). An intelligent approach to data extraction and task identification for process mining. *Information Systems Frontiers*, 17(6), 1195–1208.

Li, K., & Du, T. (2012). Building a targeted mobile advertising system for location-based services. *Decision Support Systems*, 54(1), 1–8.

Li, N., & Wu, D. (2010). Using text mining and sentiment analysis for online forums hotspot detection and forecast. *Decision Support Systems*, 48(2), 354–368.

Li, Y. M., Lin, L., & Chiu, S. W. (2014). Enhancing targeted advertising with social context endorsement. *International Journal of Electronic Commerce*, 19(1), 99–128.

Liao, J., Yang, D., Li, T., Wang, J., Qi, Q., & Zhu, X. (2014). A scalable approach for content based image retrieval in cloud datacenter. *Information Systems Frontiers*, 16(1), 129–141.

Liu, B., Kim, H., & Pennington-Gray, L. (2015). Responding to the bed bug crisis in social media. *International Journal of Hospitality Management*, 47, 76–84.

Liu, S. Q., & Mattila, A. S. (2017). Airbnb: Online targeted advertising, sense of power, and consumer decisions. *International Journal of Hospitality Management*, 60, 33–41.

Liu, X., Schuckert, M., & Law, R. (2018). Utilitarianism and knowledge growth during status seeking: Evidence from text mining of online reviews. *Tourism Management*, 66, 38–46.

Liu, Y., Teichert, T., Rossi, M., Li, H., & Hu, F. (2017). Big data for big insights: Investigating language-specific drivers of hotel satisfaction with 412,784 user-generated reviews. *Tourism Management*, 59, 554–563.

Liu, Z., Yang, P., & Zhang, L. (2013). A sketch of big data technologies. In *2013 Seventh International Conference on Internet Computing for Engineering and Science* (pp. 26–29). IEEE.

Lo, S. (2008). Web service quality control based on text mining using support vector machine. *Expert Systems with Applications*, 34(1), 603–610.

Lu, Y., Jerath, K., & Singh, P. V. (2013). The emergence of opinion leaders in a networked online community: A dyadic model with time dynamics and a heuristic for fast estimation. *Management Science*, 59(8), 1783–1799.

Ludwig, S., De Ruyter, K., Friedman, M., Brüggen, E., Wetzels, M., & Pfann, G. (2013). More than words: The influence of affective content and linguistic style matches in online reviews on conversion rates. *Journal of Marketing*, 77(1), 87–103.

Ludwig, S., De Ruyter, K., Mahr, D., Wetzels, M., Bruggen, E., & De Ruyck, T. (2014). Take their word for it: The symbolic role of linguistic style matches in user communities. *MIS Quarterly,* 38(4), 1201–1217.

Lumsdaine, A., Gregor, D., Hendrickson, B., & Berry, J. (2007). Challenges in parallel graph processing. *Parallel Processing Letters*, 17(01), 5–20.

Luo, C., Wu, F., Sun, J., & Chen, C. W. (2009). Compressive data gathering for large-scale wireless sensor networks. In *Proceedings of the 15th Annual International Conference on Mobile Computing and Networking* (pp. 145–156). ACM.

Luo, X., Andrews, M., Fang, Z., & Phang, C. W. (2014). Mobile targeting. *Management Science*, 60(7), 1738–1756.

Luo, X., Zhang, J., & Duan, W. (2013). Social media and firm equity value. *Information Systems Research*, 24(1), 146–163.

Lusch, R. F., & Nambisan, S. (2015). Service innovation: A service-dominant logic perspective. *MIS Quarterly,* 39(1), 155–175.

Lusch, R. F., Vargo, S. L., & O'Brien, M. (2007). Competing through service: Insights from service-dominant logic. *Journal of Retailing*, 83(1), 5–18.

Maglio, P. P., & Spohrer, J. (2008). Fundamentals of service science. *Journal of the Academy of Marketing Science*, 36(1), 18–20.

Magnini, V. P., Crotts, J. C., & Zehrer, A. (2011). Understanding customer delight: An application of travel blog analysis. *Journal of Travel Research*, 50(5), 535–545.

Maletic, J. I., & Marcus, A. (2000). Data cleansing: Beyond integrity analysis. In *IQ* (pp. 200–209).

Malewicz, G., Austern, M. H., Bik, A. J., Dehnert, J. C., Horn, I., Leiser, N., & Czajkowski, G. (2010). Pregel: A system for large-scale graph processing. In *Proceedings of the 2010 ACM SIGMOD International Conference on Management of Data* (pp. 135–146). ACM.

Malthouse, E. C., & Li, H. (2017). Opportunities for and Pitfalls of using big data in advertising research. *Journal of Advertising*, 46(2), 227–235.

Malthouse, E. C., Haenlein, M., Skiera, B., Wege, E., & Zhang, M. (2013). Managing customer relationships in the social media era: Introducing the social CRM house. *Journal of Interactive Marketing*, 27(4), 270–280.

Manyika, J., Chui, M., Brown, B., Bughin, J., Dobbs, R., Roxburgh, C., & Byers, A. H. (2011). *Big data: The next frontier for innovation, competition, and productivity.* NY: McKinsey Global Institute.

Mariani, M. M., Di Felice, M., & Mura, M. (2016). Facebook as a destination marketing tool: Evidence from Italian regional destination management organizations. *Tourism Management*, 54, 321–343.

Maritan, C. A., & Lee, G. K. (2017). Resource allocation and strategy. *Journal of Management*, 43(8), 2411–2420.

Mathwick, C., & Mosteller, J. (2017). Online reviewer engagement: A typology based on reviewer motivations. *Journal of Service Research*, 20(2), 204–218.

Mayer-Schönberger, V., & Cukier, K. (2013). *Big Data: A revolution that will transform how we live, work and think*. London: John Murray.

Mayzlin, D. (2006). Promotional chat on the internet. *Marketing Science*, 25(2), 155–163.

Mayzlin, D., & Yoganarasimhan, H. (2012). Link to success: How blogs build an audience by promoting rivals. *Management Science*, 58(9), 1651–1668.

McAfee, A., & Brynjolfsson, E. (2012). Big data: The management revolution, *Harvard Business Review,* 90(10), 60–68.

McCallum, A. K. (2002). Mallet: A machine learning for language toolkit, available at http://mallet.cs.umass.edu (accessed 23 November 2016).

McCreary, D., & Kelly, A. (2013). *Making sense of NoSQL*. Greenwich: Manning Publications.

McGrath, R. G. (2013). Transient advantage. *Harvard Business Review*, 91(6), 62–70.

Menon, G., Raghubir, P., & Schwarz, N. (1995). Behavioral frequency judgments: An accessibility-diagnosticity framework. *Journal of Consumer Research*, 22(2), 212–228.

Miller, A., & Tucker, C. (2013). Active social media management: The case of health care. *Information Systems Research*, 24(1), 52–70.

Min, H., & Min, H. (1997). Benchmarking the quality of hotel services: Managerial perspectives. *International Journal of Quality & Reliability Management*, 14(6), 582–597.

Min, H., Lim, Y., & Magnini, V. P. (2015). Factors affecting customer satisfaction in responses to negative online hotel reviews: The impact of empathy, paraphrasing, and speed. *Cornell Hospitality Quarterly*, 56(2), 223–231.

Min, H., Min, H., & Chung, K. (2002). Dynamic benchmarking of hotel service quality. *Journal of Services Marketing*, 16(4), 302–321.

Moe, W., & Trusov, M. (2011). The value of social dynamics in online product ratings forums. *Journal of Marketing Research*, 48(3), 444–456.

Mohsin, A., & Lockyer, T. (2010). Customer perceptions of service quality in luxury hotels in New Delhi, India: An exploratory study. *International Journal of Contemporary Hospitality Management*, 22(2), 160–173.

Moniruzzaman, A. B. M., & Hossain, S. A. (2013). NoSQL database: New era of databases for big data analytics-classification, characteristics and comparison. *International Journal of Database Theory and Application,* 6(4), 1–14.

Moon, S., Park, Y., & Kim, Y. S. (2014). The impact of text product reviews on sales. *European Journal of Marketing*, 48(11/12), 2176–2197.

Moreau, C. P., & Herd, K. B. (2009). To each his own? How comparisons with others influence consumers' evaluations of their self–designed products. *Journal of Consumer Research*, 36(5), 806–819.

Moreta, S., & Telea, A. (2007). Multiscale visualization of dynamic software logs. In *Proceedings of the 9th Joint Eurographics/IEEE VGTC Conference on Visualization* (pp. 11–18). Eurographics Association.

Morgan, N. A. (2012). Marketing and business performance. *Journal of the Academy of Marketing Science*, 40(1), 102–119.

Moro, S., Rita, P., & Vala, B. (2016). Predicting social media performance metrics and evaluation of the impact on brand building: A data mining approach. *Journal of Business Research*, 69(9), 3341–3351.

Mortenson, M. J., Doherty, N. F., & Robinson, S. (2015). Operational research from Taylorism to Terabytes: A research agenda for the analytics age. *European Journal of Operational Research*, 241(3), 583–595.

Mostafa, M. M. (2013). More than words: Social networks' text mining for consumer brand sentiments. *Expert Systems with Applications*, 40(10), 4241–4251.

Mount, M., & Martinez, M. (2014). Social media: A tool for open innovation. *California Management Review*, 56(4), 124–143.

Mudambi, S. M., & Schuff, D. (2010). What makes a helpful review? A study of customer reviews on Amazon.com. *MIS Quarterly,* 34(1), 185–200.

Nam, H., & Kannan, P. (2014). The informational value of social tagging networks. *Journal of Marketing*, 78(4), 21–40.

Nambisan, S., & Baron, R. A. (2007). Interactions in virtual customer environments: Implications for product support and customer relationship management. *Journal of Interactive Marketing*, 21(2), 42–62.

Nassirtoussi, A., Aghabozorgi, S., Wah, T., & Ngo, D. (2014). Text mining for market prediction: A systematic review. *Expert Systems with Applications*, 41(16), 7653–7670.

Nassirtoussi, A., Aghabozorgi, S., Wah, T., & Ngo, D. (2015). Text mining of news-headlines for FOREX market prediction: A multi-layer dimension reduction algorithm with semantics and sentiment. *Expert Systems with Applications*, 42(1), 306–324.

Netzer, O., Feldman, R., Goldenberg, J., & Fresko, M. (2012). Mine your own business: Market-structure surveillance through text mining. *Marketing Science*, 31(3), 521–543.

Neumeyer, L., Robbins, B., Nair, A., & Kesari, A. (2010). S4: Distributed stream computing platform. In *2010 IEEE International Conference on Data Mining Workshops* (pp. 170–177). IEEE.

Ngo-Ye, T., & Sinha, A. (2014). The influence of reviewer engagement characteristics on online review helpfulness: A text regression model. *Decision Support Systems*, 61, 47–58.

Nguyen, B., Yu, X., Melewar, T., & Chen, J. (2015). Brand innovation and social media: Knowledge acquisition from social media, market orientation, and the moderating role of social media strategic capability. *Industrial Marketing Management*, 51, 11–25.

Nicholas, D., & Huntington, P. (2003). Micro-mining and segmented log file analysis: A method for enriching the data yield from internet log files. *Journal of Information Science*, 29(5), 391–404.

Nichols, W. (2013). Advertising analytics 2.0. *Harvard Business Review*, 91(3), 60–68.

Oestreicher-Singer, G., & Sundararajan, A. (2012). The visible hand? Demand effects of recommendation networks in electronic markets. *Management Science*, 58(11), 1963–1981.

Oh, C., Roumani, Y., Nwankpa, J. K., & Hu, H. F. (2017). Beyond likes and tweets: Consumer engagement behavior and movie box office in social media. *Information & Management*, 54(1), 25–37.

Okazaki, S., & Taylor, C. (2013). Social media and international advertising: Theoretical challenges and future directions. *International Marketing Review*, 30(1), 56–71.

Oliver, R. L. (1980). A cognitive model of the antecedents and consequences of satisfaction decisions. *Journal of Marketing Research*, 17(4), 460–469.

O'Neil, C. (2016). *Weapons of math destruction: How big data increases inequality and threatens democracy.* NY: Crown Publishing Group.

Oram, A. (2014). Big data and privacy: an uneasy face-off for government to face, available at https://www.forbes.com/sites/oreillymedia/2014/03/12/big-data-and-privacy-an-uneasy-face-off-for-government-to-face/#2fca3d3e7709 (accessed 12 March 2016).

Ordenes, F. V., Theodoulidis, B., Burton, J., Gruber, T., & Zaki, M. (2014). Analyzing customer experience feedback using text mining: A linguistics-based approach. *Journal of Service Research*, 17(3), 278–295.

O'Reilly, T. (2007). What is Web 2.0: Design patterns and business models for the next generation of software. *Communications & Strategies*, (1), 17–38.

Ostrom, A. L., Parasuraman, A., Bowen, D. E., Patricio, L., & Voss, C. A. (2015). Service research priorities in a rapidly changing context. *Journal of Service Research*, 18(2), 127–159.

Özyurt, Ö., & Köse, C. (2010). Chat mining: Automatically determination of chat conversations' topic in Turkish text based chat mediums. *Expert Systems with Applications*, 37(12), 8705–8710.

Pagani, M., & Malacarne, G. (2017). Experiential engagement and active vs. passive behavior in mobile location-based social networks: The moderating role of privacy. *Journal of Interactive Marketing*, 37, 133–148.

Pal, S., Talwar, V., & Mitra, P. (2002). Web mining in soft computing framework: Relevance, state of the art and future directions. *IEEE Transactions on Neural Network*, 13(5), 1163–1177.

Panagiotopoulos, P., Barnett, J., Bigdeli, A. Z., & Sams, S. (2016). Social media in emergency management: Twitter as a tool for communicating risks to the public. *Technological Forecasting and Social Change*, 111, 86–96.

Pang, B., & Lee, L. (2008). Opinion mining and sentiment analysis. *Foundations and Trends in Information Retrieval*, 2(1–2), 1–135.

Pansari, A., & Kumar, V. (2017). Customer engagement: The construct, antecedents, and consequences. *Journal of the Academy of Marketing Science*, 45(3), 294–311.

Parasuraman, A., Berry, L. L., & Zeithaml, V. A. (1990). Guidelines for conducting service quality research. *Marketing Research*, 2(4), 34–44.

Parasuraman, A., Berry, L. L., & Zeithaml, V. A. (1991). Refinement and reassessment of the SERVQUAL scale. *Journal of Retailing*, 67(4), 420–450.

Parasuraman, A., Zeithaml, V. A., & Berry, L. L. (1985). A conceptual model of service quality and its implications for future research. *Journal of Marketing*, 49(4), 41–50.

Parasuraman, A., Zeithaml, V. A., & Berry, L. L. (1988). SERVQUAL: A multiple-item scale for measuring consumer perceptions of service quality. *Journal of Retailing*, 64(1), 12–40.

Parasuraman, A., Zeithaml, V. A., & Berry, L. L. (1994). Reassessment of expectations as a comparison standard in measuring service quality: Implications for further research. *Journal of Marketing*, 58(1), 111–124.

Parhami, B. (2006). *Introduction to parallel processing: Algorithms and architectures*. NY: Springer Science & Business Media.

Park, S. Y., & Allen, J. P. (2013). Responding to online reviews: Problem solving and engagement in hotels. *Cornell Hospitality Quarterly*, 54(1), 64–73.

Park, Y., & Chang, K. (2009). Individual and group behavior-based customer profile model for personalized product recommendation. *Expert Systems with Applications*, 36(2), 1932–1939.

Parsons, A., Zeisser, M., & Waitman, R. (1998). Organizing today for the digital marketing of tomorrow. *Journal of Interactive Marketing*, 12(1), 31–46.

Pavlou, P. A., & Dimoka, A. (2006). The nature and role of feedback text comments in online marketplaces: Implications for trust building, price premiums, and seller differentiation. *Information Systems Research*, 17(4), 392–414.

Pedersen, T., & Jensen, C. (2001). Multidimensional database technology. *Computer*, 34(12), 40–46.

Perri, T. (2002). Who wants privacy protection, and what do they want? *Journal of Consumer Behaviour*, 2(1), 80–100.

Perrons, R. K., & Jensen, J. W. (2015). Data as an asset: What the oil and gas sector can learn from other industries about "Big Data". *Energy Policy*, 81, 117–121.

Peters, K., Chen, Y., Kaplan, A. M., Ognibeni, B., & Pauwels, K. (2013). Social media metrics—A framework and guidelines for managing social media. *Journal of Interactive Marketing*, 27(4), 281–298.

Phillips, F. (2017). A Perspective on 'Big Data'. *Science and Public Policy*, 4(5), 730–737.

Poetz, M. K., & Schreier, M. (2012). The value of crowdsourcing: Can users really compete with professionals in generating new product ideas?. *Journal of Product Innovation Management*, 29(2), 245–256.

Porter, C. E., & Donthu, N. (2008). Cultivating trust and harvesting value in virtual communities. *Management Science*, 54(1), 113–128.

Prahalad, C. K., & Ramaswamy, V. (2004). Co-creation experiences: The next practice in value creation. *Journal of Interactive Marketing*, 18(3), 5–14.

Prahalad, C. K., & Ramaswamy, V. (2013). *The future of competition: Co-creating unique value with customers*. Boston: Harvard Business Press.

Proserpio, D., & Zervas, G. (2017). Online reputation management: Estimating the impact of management responses on consumer reviews. *Marketing Science*, 36(5), 645–665.

Provost, F., & Fawcett, T. (2013). Data science and its relationship to big data and data-driven decision making. *Big Data*, 1(1), 51–59.

Ramakrishnan, R., & Gehrke, J. (2000). *Database management systems*. McGraw-Hill.

Ramanathan, R. (2012). An exploratory study of marketing, physical and people related performance criteria in hotels. *International Journal of Contemporary Hospitality Management*, 24(1), 44–61.

Ramani, G., & Kumar, V. (2008). Interaction orientation and firm performance. *Journal of Marketing*, 72(1), 27–45.

Rapp, A., Beitelspacher, L. S., Grewal, D., & Hughes, D. E. (2013). Understanding social media effects across seller, retailer, and consumer interactions. *Journal of the Academy of Marketing Science*, 41(5), 547–566.

Reed, B., Chron, E., Burns, R., & Long, D. (2000). Authenticating network attached storage. *IEEE Micro*, 20(1), 49–57.

Ren, S. J., Fosso Wamba, S., Akter, S., Dubey, R., & Childe, S. J. (2016). Modelling quality dynamics, business value and firm performance in a big data analytics environment. *International Journal of Production Research*, 55(17), 5011–5026.

Rhee, H. T., & Yang, S. B. (2015). Does hotel attribute importance differ by hotel? Focusing on hotel star-classifications and customers' overall ratings. *Computers in Human Behavior*, 50, 576–587.

Rifkin, J. (2014). *The zero marginal cost society: The Internet of things, the collaborative commons, and the eclipse of capitalism.* Hampshire, UK: Palgrave Macmillan.

Risius, M., & Beck, R. (2015). Effectiveness of corporate social media activities in increasing relational outcomes. *Information & Management*, 52(7), 824–839.

Roberts, C. (2006). Radio frequency identification (RFID). *Computers & Security*, 25(1), 18–26.

Roosta, S. H. (2012). *Parallel processing and parallel algorithms: Theory and computation.* NY: Springer Science & Business Media.

Rosenbaum, P. R., & Rubin, D. B. (1985). Constructing a control group using multivariate matched sampling methods that incorporate the propensity score. *The American Statistician*, 39(1), 33–38.

Roth, P., Bobko, P., Van Iddekinge, C., & Thatcher, J. (2013). Social media in employee-selection-related decisions: A research agenda for uncharted territory. *Journal of Management*, 42(1), 269–298.

Rubin, D. B. (1973). Matching to remove bias in observational studies. *Biometrics*, 159–183.

Russom, P. (2011). Big data analytics. *TDWI Best Practices Report, Fourth Quarter*, 1–35.

Russom, P. (2012). *Analytic databases for big data.* The Data Warehousing Institute.

Rust, R. T., & Thompson, D. V. (2006). How does marketing strategy change in a service-based world. In R.F. Lusch & S.L. Vargo (Eds.), *The service-dominant logic of marketing: Dialog, debate, and directions* (pp. 381–392). NY: Routledge.

Rust, R. T., Moorman, C., & Bhalla, G. (2010). Rethinking marketing. *Harvard Business Review*, 88(1/2), 94–101.

Rust, R., & Huang, M. (2014). The service revolution and the transformation of marketing science. *Marketing Science*, 33(2), 206–221.

Ryu, G., & Feick, L. (2007). A penny for your thoughts: Referral reward programs and referral likelihood. *Journal of Marketing*, 71(1), 84–94.

Sabnis, G., & Grewal, R. (2015). Cable news wars on the internet: Competition and user-generated content. *Information Systems Research*, 26(2), 301–319.

Sagiroglu, S., & Sinanc, D. (2013). Big data: A review. In *Proceedings of the 2013 International Conference on Collaboration Technologies and Systems* (pp. 42–47). IEEE.

Saleh, F., & Ryan, C. (1991). Analysing service quality in the hospitality industry using the SERVQUAL model. *Service Industries Journal*, 11(3), 324–345.

Salihoglu, S., & Widom, J. (2013). GPS: A graph processing system. In *Proceedings of the 25th International Conference on Scientific and Statistical Database Management* (Article No. 22). ACM.

Salton, G., & McGill M. J. (1986). *Introduction to modern information retrieval.* NY: McGraw-Hill.

Sarawagi, S., & Bhamidipaty, A. (2002). Interactive deduplication using active learning. In *Proceedings of the eighth ACM SIGKDD International Conference on Knowledge Discovery and Data Mining* (pp. 269–278). ACM.

Saunders, M., Lewis, P., & Thornhill, A. (2009). *Research methods for business students* (5th ed.). Pearson.

Saunders, M., Lewis, P., & Thornhill, A. (2016). *Research methods for business students* (7th ed.). Pearson.

Schäfer, K., & Kummer, T. (2013). Determining the performance of website-based relationship marketing. *Expert Systems with Applications*, 40(18), 7571–7578.

Schneider, M. J., & Gupta, S. (2016). Forecasting sales of new and existing products using consumer reviews: A random projections approach. *International Journal of Forecasting*, 32(2), 243–256.

Schniederjans, D., Cao, E. S., & Schniederjans, M. (2013). Enhancing financial performance with social media: An impression management perspective. *Decision Support Systems*, 55(4), 911–918.

Schreier, M., & Prügl, R. (2008). Extending lead-user theory: Antecedents and consequences of consumers' lead userness. *Journal of Product Innovation Management*, 25(4), 331–346.

Schreier, M., Fuchs, C., & Dahl, D. W. (2012). The innovation effect of user design: Exploring consumers' innovation perceptions of firms selling products designed by users. *Journal of Marketing*, 76(5), 18–32.

Schweidel, D., & Moe, W. (2014). Listening in on social media: A Joint model of sentiment and venue format choice. *Journal of Marketing Research*, 51(4), 387–402.

Scullion, H., & Linehan, M. (Eds.) (2005). *International human resource management.* London: Palgrave.

Sekaran, U., & Bougie, R. (2016). *Research methods for business: A skill building approach.* Chichester: John Wiley & Sons.

Seth, N., Deshmukh, S. G., & Vrat, P. (2005). Service quality models: A review. *International Journal of Quality & Reliability Management*, 22(9), 913–949.

Shah, N., Irani, Z., & Sharif, A. M. (2017). Big data in an HR context: Exploring organizational change readiness, employee attitudes and behaviors. *Journal of Business Research*, 70, 366–378.

Shankar, V., Kleijnen, M., Ramanathan, S., Rizley, R., Holland, S., & Morrissey, S. (2016). Mobile shopper marketing: Key issues, current insights, and future research avenues. *Journal of Interactive Marketing*, 34, 37–48.

Shapiro, C. (1989). The theory of business strategy. *The RAND Journal of Economics*, 20(1), 125–137.

Sharma, S., Niedrich, R. W., & Dobbins, G. (1999). A framework for monitoring customer satisfaction: An empirical illustration. *Industrial Marketing Management*, 28(3), 231–243.

Shen, G. C. C., Chiou, J. S., Hsiao, C. H., Wang, C. H., & Li, H. N. (2016). Effective marketing communication via social networking site: The moderating role of the social tie. *Journal of Business Research*, 69(6), 2265–2270.

Sheng, J., Amankwah-Amoah, J., & Wang, X. (2017). A multidisciplinary perspective of big data in management research. *International Journal of Production Economics*, 191, 97–112.

Shieh, W. (2011). OFDM for flexible high-speed optical networks. *Journal of Lightwave Technology*, 29(10), 1560–1577.

Shiroishi, Y., Fukuda, K., Tagawa, I., Iwasaki, H., Takenoiri, S., Tanaka, H., ... & Yoshikawa, N. (2009). Future options for HDD storage. *IEEE Transactions on Magnetics*, 45(10), 3816–3822.

Shoham, M., Moldovan, S., & Steinhart, Y. (2017). Positively useless: Irrelevant negative information enhances positive impressions. *Journal of Consumer Psychology*, 27(2), 147–159.

Shriver, S. K., Nair, H. S., & Hofstetter, R. (2013). Social ties and user-generated content: Evidence from an online social network. *Management Science*, 59(6), 1425–1443.

Shvachko, K., Kuang, H., Radia, S., & Chansler, R. (2010). The Hadoop distributed file system. In *2010 IEEE 26th Symposium on Mass Storage Systems and Technologies* (pp. 1–10). IEEE.

Singh, J. (2014). Big data analytic and mining with machine learning algorithm. *International Journal of Information and Computation Technology*, 4(1), 33–40.

Singh, J. P., Irani, S., Rana, N. P., Dwivedi, Y. K., Saumya, S., & Roy, P. K. (2017). Predicting the "helpfulness" of online consumer reviews. *Journal of Business Research*, 70, 346–355.

Singh, P. V., Sahoo, N., & Mukhopadhyay, T. (2014). How to attract and retain readers in enterprise blogging?. *Information Systems Research*, 25(1), 35–52.

Singh, S., Hillmer, S., & Wang, Z. (2011). Efficient methods for sampling responses from large-scale qualitative data. *Marketing Science*, 30(3), 532–549.

Sivarajah, U., Kamal, M. M., Irani, Z., & Weerakkody, V. (2017). Critical analysis of big data challenges and analytical methods. *Journal of Business Research*, 70, 263–286.

Slevitch, L., & Oh, H. (2010). Asymmetric relationship between attribute performance and customer satisfaction: A new perspective. *International Journal of Hospitality Management*, 29(4), 559–569.

Sonnier, G. P., McAlister, L., & Rutz, O. J. (2011). A dynamic model of the effect of online communications on firm sales. *Marketing Science*, 30(4), 702–716.

Sparks, B. A., & Browning, V. (2011). The impact of online reviews on hotel booking intentions and perception of trust. *Tourism Management*, 32(6), 1310–1323.

Sparks, B. A., So, K. K. F., & Bradley, G. L. (2016). Responding to negative online reviews: The effects of hotel responses on customer inferences of trust and concern. *Tourism Management*, 53, 74–85.

Sridhar, S., & Srinivasan, R. (2012). Social influence effects in online product ratings. *Journal of Marketing*, 76(5), 70–88.

Srinivasan, S., Vanhuele, M., & Pauwels, K. (2010). Mind-set metrics in market response models: An integrative approach. *Journal of Marketing Research*, 47(4), 672–684.

Stonebraker, M., Çetintemel, U., & Zdonik, S. (2005). The 8 requirements of real-time stream processing. *ACM SIGMOD Record*, 34(4), 42–47.

Subramony, M., & Pugh, S. D. (2015). Services management research: Review, integration, and future directions. *Journal of Management*, 41(1), 349–373.

Suh, J. H., Park, C. H., & Jeon, S. H. (2010). Applying text and data mining techniques to forecasting the trend of petitions filed to e-People. *Expert Systems with Applications*, 37(10), 7255–7268.

Sun, M. (2012). How does the variance of product ratings matter?. *Management Science*, 58(4), 696–707.

Suneetha, K. R., & Krishnamoorthi, R. (2009). Identifying user behavior by analyzing web server access log file. *IJCSNS International Journal of Computer Science and Network Security*, 9(4), 327–332.

Sureshchandar, G. S., Rajendran, C., & Anantharaman, R. N. (2002). The relationship between service quality and customer satisfaction—A factor specific approach. *Journal of Services Marketing*, 16(4), 363–379.

Talia, D. (2013). Clouds for scalable big data analytics. *Computer,* 46(5), 98–101.

Tambe, P. (2014). Big data investment, skills, and firm value. *Management Science*, 60(6), 1452–1469.

Tan, K., Zhan, Y., Ji, G., Ye, F., & Chang, C. (2015). Harvesting big data to enhance supply chain innovation capabilities: An analytic infrastructure based on deduction graph. *International Journal of Production Economics*, 165, 223–233.

Tang, C., & Guo, L. (2015). Digging for gold with a simple tool: Validating text mining in studying electronic word-of-mouth (eWOM) communication. *Marketing Letters*, 26(1), 67–80.

Teece, D. J. (2007). Explicating dynamic capabilities: The nature and microfoundations of (sustainable) enterprise performance. *Strategic Management Journal*, 28(13), 1319–1350.

Teece, D. J., Pisano, G., & Shuen, A. (1997). Dynamic capabilities and strategic management. *Strategic Management Journal*, 18(7), 509–533.

Telikepalli, R., Drwiega, T., & Yan, J. (2004). Storage area network extension solutions and their performance assessment. *IEEE Communications Magazine*, 42(4), 56–63.

Thelwall, M. (2001a). A web crawler design for data mining. *Journal of Information Science*, 27(5), 319–325.

Thelwall, M. (2001b). Web log file analysis: backlinks and queries. *Aslib Proceedings*, 53(6), 217–223.

Thorleuchter, D., & Van den Poel, D. (2012). Predicting e-commerce company success by mining the text of its publicly-accessible website. *Expert Systems with Applications*, 39(17), 13026–13034.

Thorleuchter, D., & Van den Poel, D. (2013). Web mining based extraction of problem solution ideas. *Expert Systems with Applications*, 40(10), 3961–3969.

Thorleuchter, D., Van den Poel, D., & Prinzie, A. (2010). Mining ideas from textual information. *Expert Systems with Applications*, 37(10), 7182–7188.

Thorleuchter, D., Van den Poel, D., & Prinzie, A. (2012). Analyzing existing customers' websites to improve the customer acquisition process as well as the profitability prediction in B-to-B marketing. *Expert Systems with Applications*, 39(3), 2597–2605.

Tirunillai, S., & Tellis, G. J. (2012). Does chatter really matter? Dynamics of user-generated content and stock performance. *Marketing Science*, 31(2), 198–215.

Tirunillai, S., & Tellis, G. J. (2014). Mining marketing meaning from online chatter: Strategic brand analysis of big data using latent dirichlet allocation. *Journal of Marketing Research*, 51(4), 463–479.

Tripp, T. M., & Grégoire, Y. (2011). When unhappy customers strike back on the internet. *MIT Sloan Management Review*, 52(3), 37–44.

Trusov, M., Ma, L., & Jamal, Z. (2016). Crumbs of the cookie: User profiling in customer-base analysis and behavioral targeting. *Marketing Science*, 35(3), 405–426.

Tsai, T., & Lin, C. (2012). Exploring contextual redundancy in improving object-based video coding for video sensor networks surveillance. *IEEE Transactions on Multimedia*, 14(3), 669–682.

Turban, E., Sharda, R., Aronson, J. E., & King, D. (2008). *Business intelligence: A managerial approach*. Upper Saddle River, NJ: Pearson Prentice Hall.

Ulrich, K. T. (2011). *Design: Creation of artifacts in society*. Philadelphia: Pontifica Press.

Ur-Rahman, N., & Harding, J. (2012). Textual data mining for industrial knowledge management and text classification: A business oriented approach. *Expert Systems with Applications*, 39(5), 4729–4739.

Van Doorn, J., Lemon, K. N., Mittal, V., Nass, S., Pick, D., Pirner, P., & Verhoef, P. C. (2010). Customer engagement behavior: Theoretical foundations and research directions. *Journal of Service Research*, 13(3), 253–266.

Van Iddekinge, C., Lanivich, S., Roth, P., & Junco, E. (2016). Social media for selection? Validity and adverse impact potential of a Facebook-based assessment. *Journal of Management*, 42(7), 1811–1835.

Van Noort, G., & Willemsen, L. M. (2012). Online damage control: The effects of proactive versus reactive webcare interventions in consumer-generated and brand-generated platforms. *Journal of Interactive Marketing*, 26(3), 131–140.

VanMeter, R. A., Grisaffe, D. B., & Chonko, L. B. (2015). Of "Likes" and "Pins": The effects of consumers' attachment to social media. *Journal of Interactive Marketing*, 32, 70–88.

Vargo, S. L., & Lusch, R. F. (2004). Evolving to a new dominant logic for marketing. *Journal of Marketing*, 68(1), 1–17.

Vargo, S. L., & Lusch, R. F. (2008). Service-dominant logic: Continuing the evolution. *Journal of the Academy of marketing Science*, 36(1), 1–10.

Vargo, S. L., & Lusch, R. F. (2017). Service-dominant logic 2025. *International Journal of Research in Marketing*, 34(1), 46–67.

Vargo, S. L., Maglio, P. P., & Akaka, M. A. (2008). On value and value co-creation: A service systems and service logic perspective. *European Management Journal*, 26(3), 145–152.

Vasarhelyi, M., Kogan, A., & Tuttle, B. (2015). Big data in accounting: An overview. *Accounting Horizons*, 29(2), 381–396.

Verhoef, P. C., Reinartz, W. J., & Krafft, M. (2010). Customer engagement as a new perspective in customer management. *Journal of Service Research*, 13(3), 247–252.

Vermeulen, I. E., & Seegers, D. (2009). Tried and tested: The impact of online hotel reviews on consumer consideration. *Tourism Management*, 30(1), 123–127.

Vidgen, R., Shaw, S., & Grant, D. B. (2017). Management challenges in creating value from business analytics. *European Journal of Operational Research*, 261(2), 626–639.

Von Hippel E. (2005). *Democratizing innovation*. Cambridge, MA: MIT Press.

Voorhees, C. M., Brady, M. K., & Horowitz, D. M. (2006). A voice from the silent masses: An exploratory and comparative analysis of noncomplainers. *Journal of the Academy of Marketing Science*, 34(4), 514–527.

Wallach, H. M., Murray, I., Salakhutdinov, R., & Mimno, D. (2009). Evaluation methods for topic models. In *Proceedings of the 26th Annual International Conference on Machine Learning* (pp. 1105–1112), ACM.

Walther, J. B. (1992). Interpersonal effects in computer-mediated interaction: A relational perspective. *Communication Research*, 19(1), 52–90.

Walther, J. B. (1996). Computer–mediated communication: Impersonal, interpersonal, and hyperpersonal interaction. *Communication Research*, 23(1), 3–43.

Walther, J. B. (2008). Social information processing theory. In L. A. Baxter & D. O. Braithwaite (Eds.), *Engaging theories in interpersonal communication: Multiple perspectives* (pp. 391–404). California: Sage Publications.

Walther, J. B., Liang, Y. J., Ganster, T., Wohn, D. Y., & Emington, J. (2012). Online reviews, helpfulness ratings, and consumer attitudes: An extension of congruity theory to multiple sources in Web 2.0. *Journal of Computer-Mediated Communication*, 18(1), 97–112.

Fosso Wamba, S, Akter, S., Edwards, A., Chopin, G., & Gnanzou, D. (2015). How 'big data' can make big impact: Findings from a systematic review and a longitudinal case study. *International Journal of Production Economics*, 165, 234–246.

Fosso Wamba, S. F., Gunasekaran, A., Akter, S., Ren, S. J. F., Dubey, R., & Childe, S. J. (2017). Big data analytics and firm performance: Effects of dynamic capabilities. *Journal of Business Research*, 70, 356–365.

Wang, C., Lu, J., & Zhang, G. (2007). Mining key information of web pages: A method and its application. *Expert Systems with Applications*, 33(2), 425–433.

Wang, F., & Liu, J. (2011). Networked wireless sensor data collection: Issues, challenges, and approaches. *IEEE Communications Surveys & Tutorials*, 13(4), 673–687.

Wang, G., Gunasekaran, A., Ngai, E. W., & Papadopoulos, T. (2016). Big data analytics in logistics and supply chain management: Certain investigations for research and applications. *International Journal of Production Economics*, 176, 98–110.

Wang, J., Aribarg, A., & Atchadé, Y.F. (2013). Modeling choice interdependence in a social network. *Marketing Science*, 32(6), 977–997.

Wang, K. Y., Ting, I. H., & Wu, H. J. (2013). Discovering interest groups for marketing in virtual communities: An integrated approach. *Journal of Business Research*, 66(9), 1360–1366.

Wang, M., Ni, B., Hua, X. S., & Chua, T. S. (2012). Assistive tagging: A survey of multimedia tagging with human–computer joint exploration. *ACM Computing Surveys (CSUR)*, 44(4), Article No. 25.

Wang, Y., & Chaudhry, A. (2018). When and how managers' responses to online reviews affect subsequent reviews. *Journal of Marketing Research*, 55(2), 163–177.

Wang, Z., & Kim, H. G. (2017). Can social media marketing improve customer relationship capabilities and firm performance? Dynamic capability perspective. *Journal of Interactive Marketing*, 39, 15–26.

Wangenheim, F. V., & Bayón, T. (2007). Behavioral consequences of overbooking service capacity. *Journal of Marketing*, 71(4), 36–47.

Watson, H. J. (2014). Tutorial: Big data analytics: Concepts, technologies, and applications. *Communications of the Association for Information Systems*, *34*(1), 1247–1268.

Watson, H., & Wixom, B. (2007). The current state of business intelligence. *Computer*, 40(9), 96–99.

Wedel, M., & Kannan, P. K. (2016). Marketing analytics for data-rich environments. *Journal of Marketing*, 80(6), 97–121.

Wei, C., Chiang, R., & Wu, C. (2006). Accommodating individual preferences in the categorization of documents: A personalized clustering approach. *Journal of Management Information Systems*, 23(2), 173–201.

Wei, C., Hu, P., Tai, C., Huang, C., & Yang, C. (2007). Managing word mismatch problems in information retrieval: A topic-based query expansion approach. *Journal of Management Information Systems*, 24(3), 269–295.

Wei, C., Yang, C., & Hsiao, H. (2008). A collaborative filtering-based approach to personalized document clustering. *Decision Support Systems*, 45(3), 413–428.

Wei, C., Yang, C., & Lin, C. (2008). A Latent Semantic Indexing-based approach to multilingual document clustering. *Decision Support Systems*, 45(3), 606–620.

Wei, H. L., & Wang, E. T. (2010). The strategic value of supply chain visibility: Increasing the ability to reconfigure. *European Journal of Information Systems*, 19(2), 238–249.

Wei, W., Miao, L., & Huang, Z. J. (2013). Customer engagement behaviors and hotel responses. *International Journal of Hospitality Management*, 33, 316–330.

Wirtz, J., & Mattila, A. S. (2004). Consumer responses to compensation, speed of recovery and apology after a service failure. *International Journal of Service Industry Management*, 15(2), 150–166.

Witten, I. H., Frank, E., & Hall, M. A. (2011). *Data mining: Practical machine learning tools and techniques* (3rd ed.). MA: Morgan Kaufmann.

Woerner, S., & Wixom, B. (2015). Big data: Extending the business strategy toolbox. *Journal of Information Technology*, 30(1), 60–62.

Wu, L. (2013). Social network effects on productivity and job security: Evidence from the adoption of a social networking tool. *Information Systems Research*, 24(1), 30–51.

Wu, X., Kumar, V., Quinlan, J. R., Ghosh, J., Yang, Q., Motoda, H., ... & Zhou, Z. H. (2008). Top 10 algorithms in data mining. *Knowledge and Information Systems*, 14(1), 1–37.

Xiang, Z., Du, Q., Ma, Y., & Fan, W. (2017). A comparative analysis of major online review platforms: Implications for social media analytics in hospitality and tourism. *Tourism Management*, 58, 51–65.

Xiang, Z., Schwartz, Z., Gerdes, J. H., & Uysal, M. (2015). What can big data and text analytics tell us about hotel guest experience and satisfaction?. *International Journal of Hospitality Management*, 44, 120–130.

Xie, K. L., Kwok, L., & Wang, W. (2017). Monetizing managerial responses on TripAdvisor: Performance implications across hotel classes. *Cornell Hospitality Quarterly*, 58(3), 240–252.

Xie, K. L., So, K. K. F., & Wang, W. (2017). Joint effects of management responses and online reviews on hotel financial performance: A data-analytics approach. *International Journal of Hospitality Management*, 62, 101–110.

Xie, K. L., Zhang, Z., & Zhang, Z. (2014). The business value of online consumer reviews and management response to hotel performance. *International Journal of Hospitality Management,* 43, 1–12.

Xie, K. L., Zhang, Z., Zhang, Z., Singh A., & Lee, S. K. (2016). Effects of managerial response on consumer eWOM and hotel performance: Evidence from TripAdvisor. *International Journal of Contemporary Hospitality Management*, 28(9), 2013–2034.

Xu, J., Forman, C., Kim, J., & Van Ittersum, K. (2014). News media channels: Complements or substitutes? Evidence from mobile phone usage. *Journal of Marketing*, *78*(4), 97–112.

Yadav, M. S., & Pavlou, P. A. (2014). Marketing in computer-mediated environments: Research synthesis and new directions. *Journal of Marketing*, 78(1), 20–40.

Yadav, M. S., De Valck, K., Hennig-Thurau, T., Hoffman, D. L., & Spann, M. (2013). Social commerce: A contingency framework for assessing marketing potential. *Journal of Interactive Marketing*, 27(4), 311–323.

Yang, H. (2009). Automatic generation of semantically enriched web pages by a text mining approach. *Expert Systems with Applications*, 36(6), 9709–9718.

Yang, W. S., Cheng, H. C., & Dia, J. B. (2008). A location-aware recommender system for mobile shopping environments. *Expert Systems with Applications*, 34(1), 437–445.

Ye, Q., Gu, B., & Chen, W. (2010). Measuring the influence of managerial responses on subsequent online customer reviews—A natural experiment of two online travel agencies. Retrieved from http://dx.doi.org/10.2139/ssrn.1639683

Ye, Q., Zhang, Z., & Law, R. (2009). Sentiment classification of online reviews to travel destinations by supervised machine learning approaches. *Expert Systems with Applications*, 36(3), 6527–6535.

Yeh, I. C., Lien, C. H., Ting, T. M., & Liu, C. H. (2009). Applications of web mining for marketing of online bookstores. *Expert Systems with Applications*, 36(8), 11249–11256.

Yoon, J. (2012). Detecting weak signals for long-term business opportunities using text mining of Web news. *Expert Systems with Applications*, 39(16), 12543–12550.

You, Y., Vadakkepatt, G. G., & Joshi, A. M. (2015). A meta-analysis of electronic word-of-mouth elasticity. *Journal of Marketing*, 79(2), 19–39.

Yu, Y., Duan, W., & Cao, Q. (2013). The impact of social and conventional media on firm equity value: A sentiment analysis approach. *Decision Support Systems*, 55(4), 919–926.

Zhan, J., Loh, H. T., & Liu, Y. (2009). Gather customer concerns from online product reviews—A text summarization approach. *Expert Systems with Applications*, 36(2), 2107–2115.

Zhan, Y., Tan, K. H., Li, Y., & Tse, Y. K. (2016). Unlocking the power of big data in new product development. *Annals of Operations Research*, 1–19.

Zhang, M., Jansen, B. J., & Chowdhury, A. (2011). Business engagement on Twitter: A path analysis. *Electronic Markets*, 21(3), 161–175.

Zhang, X., Li, S., Burke, R. R., & Leykin, A. (2014). An examination of social influence on shopper behavior using video tracking data. J*ournal of Marketing*, 78(5), 24–41.

Zhang, Y., & Jiao, J. (2007). An associative classification-based recommendation system for personalization in B2C e-commerce applications. *Expert Systems with Applications*, 33(2), 357–367.

Zhong, R. Y., Huang, G. Q., Lan, S., Dai, Q. Y., Chen, X., & Zhang, T. (2015). A big data approach for logistics trajectory discovery from RFID-enabled production data. *International Journal of Production Economics*, 165, 260–272.

Zhou, L., Ye, S., Pearce, P. L., & Wu, M. Y. (2014). Refreshing hotel satisfaction studies by reconfiguring customer review data. *International Journal of Hospitality Management*, 38, 1–10.

Zhu, F., & Zhang, X. (2010). Impact of online consumer reviews on sales: The moderating role of product and consumer characteristics. *Journal of Marketing*, 74(2), 133–148.

Zissis, D., & Lekkas, D. (2011). Securing e-Government and e-Voting with an open cloud computing architecture. *Government Information Quarterly*, 28(2), 239–251.